PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/12, C12P 21/02, C07K 14/47, 14/705 — A1

(11) International Publication Number: **WO 00/14233**

(43) International Publication Date: 16 March 2000 (16.03.00)

(21) International Application Number: PCT/US99/20347

(22) International Filing Date: 3 September 1999 (03.09.99)

(30) Priority Data:

09/146,416	3 September 1998 (03.09.98)	US
09/390,039	3 September 1999 (03.09.99)	US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors: GLUCKSMANN, Maria, Alexandra; 33 Summit Road, Lexington, MA 02173 (US). HODGE, Martin, R.; 39 Crawford Street, Arlington, MA 02474 (US).

(74) Agents: BROWN, Anne et al.; Alston & Bird LLP, P.O. Drawer 34009, Charlotte, NC 28234-4009 (US).

(81) Designated States: AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), DM, EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: 14275 RECEPTOR, A NOVEL G-PROTEIN COUPLED RECEPTOR RELATED TO THE EDG RECEPTOR FAMILY

(57) Abstract

The present invention relates to a newly identified member of the superfamily of G-protein-coupled receptors, and a new member of the EDG receptor family. The invention also relates to polynucleotides encoding the receptor. The invention further relates to methods using receptor polypeptides and polynucleotides as a target for diagnosis and treatment in receptor-mediated disorders. The invention further relates to drug-screening methods using the receptor polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the receptor polypeptides and polynucleotides. The invention further relates to procedures for producing the receptor polypeptides and polynucleotides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

14275 RECEPTOR, A NOVEL G-PROTEIN COUPLED RECEPTOR RELATED TO THE EDG RECEPTOR FAMILY

FIELD OF THE INVENTION

The present invention relates to a newly identified member of the superfamily of G-protein-coupled receptors, and a new member of the EDG receptor family. The invention also relates to polynucleotides encoding the receptor. The invention further relates to methods using receptor polypeptides and polynucleotides as a target for diagnosis and treatment in receptor-mediated disorders. The invention further relates to drug-screening methods using the receptor polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the receptor polypeptides and polynucleotides. The invention further relates to procedures for producing the receptor polypeptides and polynucleotides.

BACKGROUND OF THE INVENTION

G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have seven transmembrane segments. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al.*, *J. Clin. Invest. 92*:1119-1125 (1993); McKusick *et al.*, *J. Med. Genet. 30*:1-26 (1993)). Specific defects in the rhodopsin gene and the V2

vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al.*, *Annu. Rev. Genet. 26*:403-424(1992)), nephrogenic diabetes insipidus (Holtzman *et al.*, *Hum. Mol. Genet. 2*:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the beta2-adrenergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem. 60*:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science 254*:1024-1026 (1991); Lin *et al.*, *Science 254*:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science 258*:597-603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science 241*:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem. 61*:1097-1129 (1992)).

There are also a small number of other proteins which present seven putative hydrophobic segments and appear to be unrelated to GPCRs; however, they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein which has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Nat'l. Acad. Sci. USA 90*:5047-5051 (1993)). The gene *frizzled* (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, *fz* has not been shown to couple to G-proteins (Vinson *et al.*, *Nature 338*:263-264 (1989)).

G proteins represent a family of heterotrimeric proteins composed of α, β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane domains. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α-subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the βγ-subunits. The GTP-bound form of the α-subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenyl

cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α-subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference.

Lipid Ligands for GPCRs

Lysophospholipids have been shown to act on distinct G-protein-coupled receptors to serve a variety of overlapping biological functions. Lysophosphatidic acid (LPA) is an extracellular phospholipid that produces multiple cellular responses including cellular proliferation, inhibition of differentiation, cell surface fibronectin binding, tumor cell invasion, chemotaxis, Cl^- mediated membrane depolarization, increased tight junction permeability, myoblast differentiation, stimulation of fibroblast chemotaxis, acute loss of gap junctional communication, platelet aggregation, smooth muscle contraction, neurotransmitter release, stress fiber formation, cell rounding, and neurite retraction, among others. See, Moolenaar, W.H. *et al.*, *Curr. Opin. Cell Biol.* 9:168-173 (1997). LPA acts through G-protein-coupled receptors to evoke the multiple cellular responses. It is generated from activated platelets and can also be generated from microvesicles shed from blood cells challenged with inflammatory stimuli. It is one of the major mitogens found in blood serum.

The N1E-115 neuronal cell line shows morphological responses to LPA. LPA induces retraction of developing neurites and rounding of the cell body, changes driven by contraction of the actomyosin system, regulated by the GTP binding protein Rho. See, Postma, *EMBO J. 15*:2388-2395 (1996).

In *Xenopus* oocytes, LPA elicits oscillatory Cl^- currents. Expression depends upon a high affinity LPA receptor having features common to members of the rhodopsin seven transmembrane receptor superfamily. An antisense oligonucleotide derived from the first 5-11 amino acids selectively inhibited expression of this receptor. See, Guo *et al.*, *Proc. Nat'l. Acad. Sci. USA 93*:14367-14372 (1996).

The intracellular biochemical signaling events that mediate the effects of LPA include stimulation of phospholipase C and consequent increases in

cytoplasmic calcium concentration, inhibition of adenyl cyclase, and activation of phosphatidylinositol-3-kinase, the Ras-Raf-MAP kinase cascade and Rho GTPase and Rho-dependent kinases. The Ras-Raf-MAP kinase and Rho pathways stimulate the transcription factors ternary complex factor and serum response factor, respectively. Ternary complex factors and serum response factors synergistically activate transcription of growth-related immediate early genes such as *c-fos* by binding to serum response element (SRE) in the promoters (Hill *et al.*, *Cell 81*:1159-1170 (1995)).

LPA receptors in fibroblasts couple to at least three distinct G-proteins: G_q, G_i, and G_{12-13}. Activation of G_q stimulates phospholipase C and consequent mobilization of intracellular calcium. Activation of G_i inhibits adenyl cyclase and stimulates the Ras-Raf-MAP kinase pathway leading to transcriptional activation mediated by ternary complex factors. Activation of G_{12-13} stimulates Rho which leads to actin-based cytoskeleton changes and transcriptional activation mediated by serum response factor. The G_i and Rho-activated pathways synergistically stimulate transcription of many growth-related genes containing serum response elements in their promoters (An, *et al.*, *J. Biol. Chem. 273*:7906-7910 (1998)).

It has been reported that serum albumin contains about a dozen as yet unidentified lipids (methanol soluble) with LPA-like biological activity. See Postma, cited above.

Sphingolipids have also been reported to be involved in cell signaling. Ceramide (N-acyl-sphingosine), sphingosine and sphingosine-1-phosphate (S1P) are second messengers involved in various biological functions. Ceramide is involved in apoptosis. S1P is a platelet-derived lysosphingolipid that acts on cognate G-protein-coupled receptors to evoke multiple cellular responses. See Moolenaar, cited above, and Meyer *et al.*, *FEBS. Lett. 410*:34-38 (1997) for a review. Typical receptor-mediated responses to S1P (and LPA) include stimulation of phospholipase C and consequent calcium mobilization, inhibition of adenylate cyclase, mitogen activated protein (MAP) kinase activation, DNA synthesis, mitogenesis and cytoskeletal changes, such as cell rounding and neurite retraction (Zondag, cited above), microfilament reorganization, cell migration, stress fiber formation, membrane depolarization, and fibroblast proliferation.

S1P has been shown to act on neuronal N1E-115 cells by means of a high affinity receptor, to remodel the actin cytoskeleton in a Rho-dependent manner.

See, Postma, *et al.*, cited above. Like LPA, S1P induces neurite retraction and cell rounding in differentiated PC12 cells. See, Sato *et al.*, *Biochem. Biophys. Res. Comm. 240*:329-334 (1997).

S1P acts by activating a G-protein-coupled receptor distinct from the LPA receptor.

A distinct receptor is also activated by another lysosphingolipid, sphingosyl-phosphorylcholine (SPC or lysosphingomyelin). It is a strong mitogen and evokes biochemical responses similar to those by LPA, except by a distinct receptor (in some cells, however, SPC and S1P might act on the same receptor). See, Moolenaar, cited above. SPC has also been shown to mediate fibroblast mitogenesis, platelet activation, and neurite retraction. It has been shown to activate MAP kinases. See, An *et al.*, *FEBS Lett. 417*:279-282 (1997). S1P and SPC also activate pathways involving G_i, Ras-Raf-ERK and Rho GTPases (An *et al.*, *FEBS Lett.*).

Since S1P and LPA are both released from activated platelets, they may play a role in wound healing and tissue remodeling, including during traumatic injury of the nervous system. Because LPA can also be generated from blood cells challenged with inflammatory stimuli, LPA may stimulate responses not only at the site of injury but also at sites of inflammation.

EDG receptors

Hecht *et al.* (*J. Cell Biol. 135*:1071-1083 (1996)) cloned a cDNA from mouse neocortical cell lines. This gene, termed ventricular zone gene-1 (*vzg-1*) was shown to be 96% identical to an unpublished sheep sequence designated EDG-2 (GenBank Accession No. U18405) and identified as an LPA receptor. This cDNA was also isolated as an orphan receptor by Macrae *et al.* (*Mol. Brain Res. 42*:245-254 (1996)) who designated it Rec1.3. EDG-2 is closely homologous to a G_i-linked orphan receptor EDG-1 (37% homology). A cDNA homologous to that encoding sheep EDG-2 protein was cloned from a human lung cDNA library (An *et al.*, *Biochem. Biophys. Res. Comm. 231*:619-622 (1997)). A search of GenBank showed that EDG-2 cDNA from mouse and cow had also been cloned and sequenced. The human EDG-2 protein was shown to be a receptor for LPA. The cDNA was expressed in mammalian cells (HEK293 and CHO) using a reporter gene assay quantifying the transcriptional activation of a serum response

element-containing promoter. This assay can sensitively measure the G-protein-activated signaling pathways linked to LPA receptors. The mouse EDG-2 (*Vzg-1*) showed 96% identity to the human EDG-2 (Hecht *et al.*, *J. Cell Biol. 135*:1071-1083 (1996)). EDG-2 was demonstrated to mediate inhibition of adenyl cyclase by Gi and cell morphological changes via Rho-related GTPases (An *et al.*, *J. Biol. Chem. 273*:7906-7910 (1998)).

Human EDG-1 cDNA was cloned from a human cDNA library of human umbilical vein endothelial cells exposed to fluid sheer stress (Takada *et al.*, *Biochem. Biophys. Res. Comm. 240*:737-741 (1997)). EDG-1 mRNA levels in endothelial cells increased markedly in response to fluid flow. This suggested that EDG-1 is a receptor gene that could function to regulate endothelial function under physiological blood flow conditions. Recently, it was shown that the EDG-1 receptor is capable of mediating a subset of early responses to sphingosine 1-phosphate (S1P), notably, inhibition of adenylate cyclase and activation of the G_i-MAP kinase pathway, but not activation of the PLC-Ca^{2+} signaling pathway. (Zondag, G.C. *et al.*, *Bio. Chem. J. 330*:605-609 (1998)).

In the study of Zondag, the results indicated that EDG-1 but not EDG-2 was capable of mediating the specific subset of cellular actions induced by S1P. However, these responses were specific in that LPA failed to mimic S1P.

Another study (Fukushima *et al.*, *Proc. Nat'l. Acad. Sci. USA 95*:6151-6156 (1998)) showed that the human EDG-2 mediates multiple cellular responses to LPA. At least six biological responses to LPA were reported, including the production of LPA membrane binding sites, LPA dependent G-protein activation, stress fiber formation, neurite retraction, transcriptional serum response element activation and increased DNA synthesis. EDG-1 and EDG-2 were shown to signal through at least two distinct pathways, a G_i/G_o pathway and a PTX insensitive pathway that involves Rho activation. It was demonstrated that G_i coupled directly with *Vzg-1* (EDG-2) after LPA exposure. At the same time it was shown that *Vzg-1* mediates actin-based cytoskeletal changes that operate through a Rho-sensitive pathway. See Fukushima, cited above. The results were consistent with a model in which EDG-2 transduces LPA signals onto the same DNA target through two separate pathways. Activation of serum response element-dependent transcription can be effected through stimulation of the Ras-Raf-MAP kinase cascade (by a ternary complex factor) and through a Rho-

mediated pathway. An important response related to the serum response element activation is progression through the cell cycle.

Using the cDNA sequence of the EDG-2 human LPA receptor to perform a sequence-based search for lysosphingolipid receptors, An *et al.* (*FEBS. Lett.* *417*:279-282 (1997)) found two closely related G-protein-coupled receptors, designated rat H218 and human EDG-3. Both of these, when overexpressed in Jurkat cells, mobilized calcium and activated serum response element-driven transcriptional reporter gene (which requires activation of Rho and Ras GTPases) in response to S1P, dihydro-S1P, and sphingosylphosphorylcholine, but not to LPA. Expressed in *Xenopus* oocytes, the genes conferred responsiveness to S1P in agonist-triggered calcium efflux.

EDG-2 was also used for a sequence-based search for new genes encoding novel subtypes of LPA receptors. A human cDNA encoding a G-protein-coupled receptor designated EDG-4 was identified by searching GenBank for homologies with the EDG-2 LPA receptor. When overexpressed in Jurkat cells, this protein mediates LPA-induced activation of a serum response element reporter gene with LPA concentration-dependence and specificity (An *et al.*, *J. Biol. Chem.* 273:7906-7910 (1998)). Jurkat cells are a preferred assay system because they lack background responses to LPA in the serum response element reporter gene assay. EDG4 was shown to mediate activation of serum response element-driven transcription in Jurkat cells involving G_i and Rho GTPase.

A flow chart designating homologies of the various EDG receptors is shown in Figure 5, *infra*.

GPCRs in general and EDG receptors are important targets for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs, particularly EDG receptors. The present invention advances the state of the art by providing a previously unidentified human GPCR, a new member of the EDG receptor family.

SUMMARY OF THE INVENTION

It is an object of the invention to identify novel GPCRs.

It is a further object of the invention to provide novel GPCR polypeptides that are useful as reagents or targets in receptor assays applicable to treatment and diagnosis of GPCR-mediated disorders.

It is a further object of the invention to provide polynucleotides corresponding to the novel GPCR polypeptides that are useful as targets and reagents in receptor assays applicable to treatment and diagnosis of GPCR-mediated disorders and useful for producing novel receptor polypeptides by recombinant methods.

A specific object of the invention is to identify compounds that act as agonists and antagonists and modulate the expression of the receptor.

A further specific object of the invention is to provide the compounds that modulate the expression of the receptor for treatment and diagnosis of GPCR related disorders.

The invention is thus based on the identification of a novel GPCR, designated the 14275 receptor.

The invention provides isolated 14275 receptor polypeptides including a polypeptide having the amino acid sequence shown in SEQ ID NO 1, or the amino acid sequence encoded by the cDNA deposited as ATCC No. _____ on _____ ("the deposited cDNA").

The invention also provides isolated 14275 receptor nucleic acid molecules having the sequence shown in SEQ ID NO 2 or in the deposited cDNA.

The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NO 1 or encoded by the deposited cDNA.

The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NO 2 or in the deposited cDNA.

The invention also provides fragments of the polypeptide shown in SEQ ID NO 1 and nucleotide shown in SEQ ID NO 2, as well as substantially homologous fragments of the polypeptide or nucleic acid.

The invention further provides nucleic acid constructs comprising the nucleic acid molecules described herein. In a preferred embodiment, the nucleic acid molecules of the invention are operatively linked to a regulatory sequence.

The invention also provides vectors and host cells for expression of the receptor nucleic acid molecules and polypeptides and particularly recombinant vectors and host cells.

The invention also provides methods of making the vectors and host cells and methods for using them to produce the receptor nucleic acid molecules and polypeptides.

The invention also provides antibodies or antigen-binding fragments thereof that selectively bind the receptor polypeptides and fragments.

The invention also provides methods of screening for compounds that modulate expression or activity of the polypeptides or nucleic acid (RNA or DNA).

The invention also provides a process for modulating polypeptide or nucleic acid expression or activity, especially using the screened compounds. Modulation may be used to treat conditions related to aberrant activity or expression of the polypeptides or nucleic acids.

The invention also provides assays for determining the presence or absence of and level of the polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

The invention also provides assays for determining the presence of a mutation in the polypeptides or nucleic acid molecules, including for disease diagnosis.

In still a further embodiment, the invention provides a computer readable means containing the nucleotide and/or amino acid sequences of the nucleic acids and polypeptides of the invention, respectively.

The invention also provides methods of screening for compounds that modulate the activity of the receptor polypeptides. Modulation can be at the level of the polypeptide receptor or at the level of controlling the expression of nucleic acid expressing the receptor polypeptide.

The invention also provides a process for modulating receptor polypeptide activity, especially using the screened compounds, including to treat conditions related to expression of the receptor polypeptides.

The invention also provides diagnostic assays for determining the presence of and level of the receptor polypeptides or nucleic acid molecules in a biological sample.

The invention also provides diagnostic assays for determining the presence of a mutation in the receptor polypeptides or nucleic acid molecules.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows the 14275 nucleotide sequence (SEQ ID NO 2) and the deduced 14275 amino acid sequence (SEQ ID NO 1). It is predicted that amino acids 1-50 constitute the amino terminal extracellular domain and amino acids 332-384 constitute the carboxy terminal intracellular domain. The region spanning the entire transmembrane domain is from about amino acid 51 to about amino acid 331. Specifically, the seven transmembrane segments are as follows: from about amino acid 51 to about amino acid 71, from about amino acid 81 to about amino acid 105, from about amino acid 123 to about amino acid 141, from about amino acid 162 to about amino acid 184, from about amino acid 204 to about amino acid 227, from about amino acid 253 to about amino acid 276, and from about amino acid 291 to about amino acid 331. The amino acids corresponding to the three extracellular loops are as follows: from about amino acid 106 to about amino acid 122, from about amino acid 185 to about amino acid 203, from about amino acid 277 to about amino acid 290. The intracellular loops are from about amino acid 72 to about amino acid 80, from about amino acid 142 to about amino acid 161, and from about amino acid 228 to about amino acid 252. The underlined area shows a GPCR signature. The most commonly conserved intracellular sequence is the aspartate, arginine, tyrosine (DRY) triplet. Arginine is invariant. Aspartate is conservatively placed in several GPCRs. DRY is implicated in signal transduction. In the present case, the arginine is found in the sequence ERF, which matches the position of the DRY or invariant arginine for a rhodopsin family seven transmembrane receptor. See Figure 6.

Figure 2 shows an analysis of the 14275 open reading frame for amino acids corresponding to specific functional sites. A glycosylation site is found at amino acids 2 to 5, which is in the amino terminal extracellular domain. A second

glycosylation site is found at amino acids 30 to 33, which is also in the amino terminal extracellular domain. A third glycosylation site is found at amino acids 87 to 90, which is in the second transmembrane segment. A protein kinase C phosphorylation site is found at amino acids 32 to 34, which is in the amino terminal extracellular domain. A second protein kinase C phosphorylation site is found at amino acids 77 to 79, which is in the first intracellular loop. A third protein kinase C phosphorylation site is found at amino acids 110 to 112, which is in the first extracellular loop. A fourth protein kinase C phosphorylation site is found at amino acids 159 to 161, which is in the second intracellular loop. A fifth protein kinase C phosphorylation site is found at amino acids 201 to 203, which is in the second extracellular loop. A sixth protein kinase C phosphorylation site is found at amino acids 308 to 310, which is in the seventh transmembrane segment. A seventh protein kinase C phosphorylation site is found at amino acids 354 to 356, which is in the carboxy terminal intracellular domain. An eighth protein kinase C phosphorylation site is found at amino acids 360 to 362, which is in the carboxy terminal intracellular domain. A ninth protein kinase C phosphorylation site is found at amino acids 368 to 370, which is in the carboxy terminal intracellular domain. A tenth protein kinase C phosphorylation site is found at amino acids 380 to 382, which is in the carboxy terminal intracellular domain. A casein kinase II phosphorylation site is found at amino acids 89 to 92, which is in the second transmembrane segment. A second casein kinase II phosphorylation site is found at amino acids 139 to 142, which spans the third transmembrane segment and second intracellular loop. A third casein kinase II phosphorylation site is found at amino acids 349 to 352, which is in the carboxy terminal intracellular domain. An N-myristoylation site is found at amino acids 44 to 49, which is in the amino terminal extracellular domain. A second N-myristoylation site is found at amino acids 51 to 56, which is in the first transmembrane segment. A third N-myristoylation site is found at amino acids 123 to 128, which is in the third transmembrane segment. A fourth N-myristoylation site is found at amino acids 155 to 160, which is in the second intracellular loop. A fifth N-myristoylation site is found at amino acids 214 to 219, which is in the fifth transmembrane segment. A sixth N-myristoylation site is found at amino acids 221 to 226, which is in the fifth transmembrane segment. A seventh N-myristoylation site is found at amino acids 269 to 274, which is in the sixth

transmembrane segment. An eighth N-myristoylation site is found at amino acids 347 to 352, which is in the carboxy terminal intracellular domain. In addition, amino acids corresponding in position to the GPCR signature and containing the invariant arginine are found in the sequence ERF at amino acids 142 to 144.

Figure 3 shows an analysis of the 14275 amino acid sequence: αβturn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability.

Figure 4 shows a 14275 receptor hydrophobicity plot. Amino acids 51-331 constitute the entire transmembrane domain that includes the seven transmembrane segments, the three intracellular loops and the three extracellular loops.

Figure 5 shows the approximate percent identity among various EDG family members as follows:

EDG1-EDG2:40%; EDG1-EDG4:40%; EDG1-EDG3:55%; EDG1-14275 receptor:43%;

EDG2-EDG4:57%; EDG2-EDG3:39%; EDG2-14275 receptor:37%;

EDG3-EDG4:32%; EDG3-14275 receptor:42%;

EDG4-14275 receptor:40%

Figure 6 shows a sequence comparison between a seven transmembrane receptor member of the rhodopsin superfamily and the 14275 receptor showing the position of the ERF, that corresponds to the GPCR signature.

DETAILED DESCRIPTION OF THE INVENTION

Receptor function/signal pathway

The 14275 receptor protein is a GPCR that participates in signaling pathways. As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR (14275 protein). Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction

pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP_2), inositol 1,4,5-triphosphate (IP_3) or adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival. Functions mediated by EDG receptors are further presented in the background section, *supra*.

Since the 14275 receptor protein is expressed in peripheral blood cells, spleen, lung, small intestine, prostate, heart, thymus, colon, uterus and placenta, cells participating in a 14275 receptor protein signaling pathway include, but are not limited to cells derived from these tissues.

Depending on the type of cell, the response mediated by the receptor protein may be different. For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the receptor protein, it is universal that the protein is a GPCR and interact with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell.

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP_2) as well as to the activities of these molecules. PIP_2 is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP_2 to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP_3). Once formed IP_3 can diffuse to the endoplasmic reticulum surface where it can bind an IP_3 receptor, e.g., a calcium channel protein containing an IP_3 binding site. IP_3 binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP_3 can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP_4), a molecule which can cause calcium entry into the cytoplasm from the extracellular medium. IP_3 and IP_4 can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-biphosphate (IP_2) and inositol 1,3,4-triphosphate,

respectively. These inactive products can be recycled by the cell to synthesize PIP_2. The other second messenger produced by the hydrolysis of PIP_2, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-kB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP_2 or one of its metabolites.

Another signaling pathway the receptor may participate in is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

Polypeptides

The invention is based on the identification of a novel G-coupled protein receptor. Specifically, an expressed sequence tag (EST) was selected based on homology to G-protein-coupled receptor sequences. This EST was used to design primers based on sequences that it contains and used to identify a cDNA from a human prostate cDNA library. Positive clones were sequenced and the overlapping fragments were assembled. Analysis of the assembled sequence revealed that the cloned cDNA molecule encodes a G-protein coupled receptor

showing a high homology score against the seven transmembrane segment rhodopsin superfamily, also with high homology to the EDG receptor family.

The invention thus relates to a novel GPCR having the deduced amino acid sequence shown in Figure 1 (SEQ ID NO 1) or having the amino acid sequence encoded by the deposited cDNA, ATCC No. _____.

The deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms. The deposit is provided as a convenience to those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112. The deposited sequence, as well as the polypeptide encoded by the sequence, is incorporated herein by reference and controls in the event of any conflict, such as a sequencing error, with description in this application.

The "14275 receptor polypeptide" or "14275 receptor protein" refers to the polypeptide in SEQ ID NO 1 or encoded by the deposited cDNA. The term "receptor protein" or "receptor polypeptide", however, further includes the numerous variants described herein, as well as fragments derived from the full length 14275 polypeptide and variants.

The present invention thus provides an isolated or purified 14275 receptor polypeptide and variants and fragments thereof.

The 14275 polypeptide is a 384 residue protein exhibiting three main structural domains. The amino terminal extracellular domain is identified to be within residues 1 to about 50 in SEQ ID NO 1. The region spanning the entire transmembrane domain is identified to be within residues from about 51 to about 331 in SEQ ID NO 1. Discrete transmembrane segments are estimated to be from about amino acid 51-71, 81-105, 123-141, 162-184, 204-227, 253-276, and 291-331. Accordingly, the six extracellular and intracellular loops correspond to about amino acids 106-122, 185-203, 277-290, 72-80, 142-161, and 228-252. The carboxy terminal intracellular domain is identified to be within residues from about 332 to about 384 in SEQ ID NO 1. The transmembrane domain includes the invariant arginine of a GPCR signal transduction signature, ERF, at residues 142-144.

A glycosylation site is found from amino acids 2 to 5, which is in the amino terminal extracellular domain. A second glycosylation site is found at amino acids 30 to 33, which is also in the amino terminal extracellular domain. A

third glycosylation site is found at amino acids 87 to 90, which is in the second transmembrane segment. A protein kinase C phosphorylation site is found at amino acids 32 to 34, which is in the amino terminal extracellular domain. A second protein kinase C phosphorylation site is found at amino acids 77 to 79, which is in the first intracellular loop. A third protein kinase C phosphorylation site is found at amino acids 110 to 112, which is in the first extracellular loop. A fourth protein kinase C phosphorylation site is found at amino acids 159 to 161, which is in the second intracellular loop. A fifth protein kinase C phosphorylation site is found at amino acids 201 to 203, which is in the second extracellular loop. A sixth protein kinase C phosphorylation site is found at amino acids 308 to 310, which is in the seventh transmembrane segment. A seventh protein kinase C phosphorylation site is found at amino acids 354 to 356, which is in the carboxy terminal intracellular domain. An eighth protein kinase C phosphorylation site is found at amino acids 360 to 362, which is in the carboxy terminal intracellular domain. A ninth protein kinase C phosphorylation site is found at amino acids 368 to 370, which is in the carboxy terminal intracellular domain. A tenth protein kinase C phosphorylation site is found at amino acids 380 to 382, which is in the carboxy terminal intracellular domain. A casein kinase II phosphorylation site is found at amino acids 89 to 92, which is in the second transmembrane segment. A second casein kinase II phosphorylation site is found at amino acids 139 to 142, which spans the third transmembrane segment and second intracellular loop. A third casein kinase II phosphorylation site is found at amino acids 349 to 352, which is in the carboxy terminal intracellular domain. An N-myristoylation site is found at amino acids 44 to 49, which is in the amino terminal extracellular domain. A second N-myristoylation site is found at amino acids 51 to 56, which is in the first transmembrane segment. A third N-myristoylation site is found at amino acids 123 to 128, which is in the third transmembrane segment. A fourth N-myristoylation site is found at amino acids 155 to 160, which is in the second intracellular loop. A fifth N-myristoylation site is found at amino acids 214 to 219, which is in the fifth transmembrane segment. A sixth N-myristoylation site is found at amino acids 221 to 226, which is in the fifth transmembrane segment. A seventh N-myristoylation site is found at amino acids 269 to 274, which is in the sixth transmembrane segment. An eighth N-myristoylation site is found at amino acids 347 to 352, which is in the carboxy terminal intracellular domain. In

addition, amino acids corresponding in position to the GPCR signature and containing the invariant arginine are found in the sequence ERF at amino acids 142 to 144.

The 14275 amino acid sequence showed approximately 40% identity with EDG-4, 37% identity with EDG-2, 42% identity with EDG-3, and 43% identity with EDG-1.

As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell and still be considered "isolated" or "purified."

The receptor polypeptides can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful and considered to contain an isolated form of the polypeptide. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity.

In one embodiment, the language "substantially free of cellular material" includes preparations of the receptor polypeptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the receptor polypeptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the protein preparation.

A polypeptide is also considered to be isolated when it is part of a membrane preparation or is purified and then reconstituted with membrane vesicles or liposomes.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the receptor polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having

less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, the receptor polypeptide comprises the amino acid sequence shown in SEQ ID NO 1. However, the invention also encompasses sequence variants. Variants include a substantially homologous protein encoded by the same genetic locus in an organism, i.e., an allelic variant. In the present case, the 14725 receptor gene has been mapped to chromosome 2, near the marker NRB733. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to the 14275 receptor protein of SEQ ID NO 1. Variants also include proteins substantially homologous to the 14275 receptor protein but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to the 14275 receptor protein that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the 14275 receptor protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

As used herein, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences are at least about 55-60%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NO 2 under stringent conditions as more fully described below.

To determine the percent homology of two amino acid sequences, or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence, then the molecules are homologous at that position. As used herein,

amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity". The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., per cent homology equals the number of identical positions/total number of positions times 100).

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by the 14275 polypeptide. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al., Science 247*:1306-1310 (1990).

TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine
	Tryptophan
	Tyrosine
Hydrophobic	Leucine
	Isoleucine
	Valine
Polar	Glutamine
	Asparagine
Basic	Arginine
	Lysine
	Histidine
Acidic	Aspartic Acid
	Glutamic Acid
Small	Alanine
	Serine
	Threonine
	Methionine
	Glycine

Both identity and similarity can be readily calculated *(Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton

Press, New York, 1991). Preferred computer program methods to determine identify and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Res. 12(1)*:387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F. *et al.*, *J. Molec. Biol. 215*:403 (1990)).

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (, % identity = # of identical positions/total # of positions (*e.g.*, overlapping positions) x 100). In one embodiment, the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast

programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov.

Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the CGC sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti (1994) *Comput. Appl. Biosci., 10*:3-5; and FASTA described in Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. 85*:2444-8. Within FASTA, ktup is a control option that sets the sensitivity and speed of the search. If ktup=2, similar regions in the two sequences being compared are found by looking at pairs of aligned residues; if ktup=1, single aligned amino acids are examined. ktup can be set to 2 or 1 for protein sequences, or from 1 to 6 for DNA sequences. The default if ktup is not specified is 2 for proteins and 6 for DNA. For a further description of FASTA parameters, see http://bioweb.pasteur.fr/docs/man/man/fasta.1.html#sect2, the contents of which are incorporated herein by reference.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these.

Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variations can affect the function, for example, of one or more of the regions corresponding to ligand binding, membrane association, G-protein binding and signal transduction.

Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids which result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

As indicated, variants can be naturally-occurring or can be made by recombinant means or chemical synthesis to provide useful and novel characteristics for the receptor polypeptide. This includes preventing immunogenicity from pharmaceutical formulations by preventing protein aggregation.

Useful variations further include alteration of ligand binding characteristics. For example, one embodiment involves a variation at the binding site that results in binding but not release of ligand. A further useful variation at the same sites can result in a higher affinity for ligand. Useful variations also include changes that provide for affinity for another ligand. Another useful variation includes one that allows binding but which prevents activation by the ligand. Another useful variation includes variation in the transmembrane G-protein-binding/signal transduction domain that provides for reduced or increased binding by the appropriate G-protein or for binding by a different G-protein than the one with which the receptor is normally associated. Another useful variation provides a fusion protein in which one or more domains or sub-regions is operationally fused to one or more domains or sub-regions from another G-protein coupled receptor.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro*, *or in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

The invention also includes polypeptide fragments of the 14275 receptor protein. Fragments can be derived from the amino acid sequence shown in SEQ ID NO 1. However, the invention also encompasses fragments of the variants of the 14275 receptor protein as described herein.

The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed prior to the present invention.

Fragments can retain one or more of the biological activities of the protein, for example the ability to bind to a G-protein or ligand. Fragments can also be useful as an immunogen to generate receptor antibodies.

Biologically active fragments can comprise a domain or motif, e.g., an extracellular or intracellular domain or loop, one or more transmembrane segments, or parts thereof, G-protein binding site, or GPCR signature, glycosylation sites, protein kinase C, or casein kinase II phosphorylation sites, and myristoylation sites. Such peptides can be, for example, 7, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length. Such domains or motifs can be identified by means of routine computerized homology searching procedures.

Possible fragments include, but are not limited to: 1) soluble peptides comprising the entire amino terminal extracellular domain about amino acid 1 to about amino acid 50 of SEQ ID NO 1, or parts thereof; 2) peptides comprising the entire carboxy terminal intracellular domain from about amino acid 332 to amino acid 384 of SEQ ID NO 1, or parts thereof; 3) peptides comprising the region spanning the entire transmembrane domain from about amino acid 51 to about amino acid 331, or parts thereof; 4) any of the specific transmembrane segments, or parts thereof, from about amino acid 51 to about amino acid 71, from about amino acid 81 to about amino acid 105, from about amino acid 123 to about amino acid 141, from about amino acid 162 to about amino acid 184, from about amino acid 204 to about amino acid 227, from about amino acid 253 to about amino acid 276, and from about amino acid 291 to about amino acid 331; 5) any of the three intracellular or three extracellular loops, or parts thereof, from about amino acid 72 to about amino acid 80, from about amino acid 142 to about amino acid 161, from about amino acid 228 to about amino acid 252, from about amino acid 106 to about amino acid 122, from about amino acid 185 to about amino acid 203, and from about amino acid 277 to about amino acid 290. Fragments further include combinations of the above fragments, such as an amino terminal domain combined with one or more transmembrane segments and the attendant extra or intracellular loops or one or more transmembrane segments, and the attendant intra or extracellular loops, plus the carboxy terminal domain. Thus, any of the

above fragments can be combined. Other fragments include the mature protein from about amino acid 6 to 384. Other fragments contain the various functional sites described herein, such as phosphorylation sites, glycosylation sites, and myristoylation sites and a sequence containing the GPCR signature sequence. Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific domains mentioned above, which sub-fragments retain the function of the domain from which they are derived. Fragments also include amino acid sequences greater than 17 amino acids. Fragments also include antigenic fragments and specifically those shown to have a high antigenic index in Figure 3. Further specific fragments include a fragment from about 1 to 120 and sub-fragments thereof greater than 6 amino acids, from about 116 to 296 and sub-fragments thereof grater than 9 amino acids, from about 288 to 361 and sub-fragments thereof greater than 10 amino acids, from about 352-384 and subfragments thereof greater than 17 amino acids, and from about 375 to 384 and sub-fragments thereof. Further fragments include a fragment including any amino acid sequences from 1-120 but extending beyond amino acid 120.

Accordingly, possible fragments include fragments defining a ligand-binding site, fragments defining a glycosylation site, fragments defining membrane association, fragments defining phosphorylation sites, fragments defining interaction with G proteins and signal transduction, and fragments defining myristoylation sites. By this is intended a discrete fragment that provides the relevant function or allows the relevant function to be identified. In a preferred embodiment, the fragment contains the ligand-binding site. These regions can be identified by well-known methods involving computerized homology analysis.

The invention also provides fragments with immunogenic properties. These contain an epitope-bearing portion of the 14275 receptor protein and variants. These epitope-bearing peptides are useful to raise antibodies that bind specifically to a receptor polypeptide or region or fragment. These peptides can contain at least 6, 9, 12, 14, or between at least about 15 to about 30 amino acids.

Non-limiting examples of antigenic polypeptides that can be used to generate antibodies include peptides derived from the amino terminal extracellular

domain or any of the extracellular loops. However, other peptides are possible, for example, intracellular regions that could serve as an intrabody target. Regions having a high antigenicity index are shown in Figure 3.

The receptor polypeptides (including variants and fragments which may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR related conditions.

The epitope-bearing receptor and polypeptides may be produced by any conventional means (Houghten, R.A., *Proc. Nat'l. Acad. Sci. USA 82*:5131-5135 (1985)). Simultaneous multiple peptide synthesis is described in U.S. Patent No. 4,631,211.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the receptor fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion proteins. These comprise a receptor protein operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the receptor protein. "Operatively linked" indicates that the receptor protein and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the receptor protein.

In one embodiment the fusion protein does not affect receptor function per se. For example, the fusion protein can be a GST-fusion protein in which the receptor sequences are fused to the C-terminus of the GST sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant receptor protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Therefore, in another

embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus.

EP-A-O 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. Bennett *et al.* (*Journal of Molecular Recognition 8*:52-58 (1995)) and Johanson *et al.* (*Journal of Biological Chemistry 270 (16)*:9459-9471 (1995)). Thus, this invention also encompasses soluble fusion proteins containing a receptor polypeptide and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE). Preferred as immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. For some uses it is desirable to remove the Fc after the fusion protein has been used for its intended purpose, for example when the fusion protein is to be used as antigen for immunizations. In a particular embodiment, the Fc part can be removed in a simple way by a cleavage sequence which is also incorporated and can be cleaved with factor Xa.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A receptor protein-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the receptor protein.

Another form of fusion protein is one that directly affects receptor functions. Accordingly, a receptor polypeptide encompassed by the present invention in which one or more of the receptor domains (or parts thereof) has been

replaced by homologous domains (or parts thereof) from another G-protein coupled receptor or other type of receptor. Accordingly, various permutations are possible. The amino terminal extracellular domain, or subregion thereof, (for example, ligand-binding) may be replaced with the domain or subregion from another ligand-binding receptor protein. Alternatively, the region spanning the entire transmembrane domain or any of the seven segments or loops, for example, G-protein-binding/signal transduction, may be replaced. Finally, the carboxy terminal intracellular domain or sub-region may be replaced. Thus, chimeric receptors can be formed in which one or more of the native domains or subregions has been replaced.

The isolated receptor protein can be purified from cells that naturally express it, such as from peripheral blood cells, such as T or B cells, mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, $CD4^+$T lymphocytes, spleen, lung, thymus, uterus, small intestine, colon, heart, prostate, and placenta, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the receptor polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.*, *Meth. Enzymol. 182*: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci. 663*:48-62 (1992).

As is also well known, polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched

circular polypeptides may be synthesized by non-translational natural processes and by synthetic methods.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

The same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain more than one type of modification.

Polypeptide uses

The receptor polypeptides are useful for producing antibodies specific for the 14275 receptor protein, regions, or fragments. Regions having a high antigenicity index are shown in Figure 3.

The receptor polypeptides (including variants and fragments which may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR-related conditions, especially disorders involving the tissues in which the receptor is expressed, such as disclosed herein.

The receptor polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native i.e., cells that

normally express the receptor protein, as a biopsy or expanded in cell culture. In one embodiment, however, cell-based assays involve recombinant host cells expressing the receptor protein.

Determining the ability of the test compound to interact with the polypeptide can also comprise determining the ability of the test compound to preferentially bind to the polypeptide as compared to the ability of the ligand, or a biologically active portion thereof, to bind to the polypeptide.

The polypeptides can be used to identify compounds that modulate receptor activity. Such compounds, for example, can increase or decrease affinity or rate of binding to a known ligand, compete with ligand for binding to the receptor, or displace ligand bound to the receptor. Both 14275 protein and appropriate variants and fragments can be used in high throughput screens to assay candidate compounds for the ability to bind to the receptor. These compounds can be further screened against a functional receptor to determine the effect of the compound on the receptor activity. Compounds can be identified that activate (agonist) or inactivate (antagonist) the receptor to a desired degree. Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject).

The receptor polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the receptor protein and a target molecule that normally interacts with the receptor protein. The target can be ligand or a component of the signal pathway with which the receptor protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). The assay includes the steps of combining the receptor protein with a candidate compound under conditions that allow the receptor protein or fragment to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the receptor protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cyclic AMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Determining the ability of the protein to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction

Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem. 63*:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol. 5*:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des. 12*:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA 90*:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA 91*:11422; Zuckermann *et al.* (1994). *J. Med. Chem. 37*:2678; Cho *et al.* (1993) *Science 261*:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl. 33*:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl. 33*:2061; and in Gallop *et al.* (1994) *J. Med. Chem. 37*:1233. Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques 13*:412-421), or on beads (Lam (1991) *Nature 354*:82-84), chips (Fodor (1993) *Nature 364*:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA 89*:1865-1869) or on phage (Scott and Smith (1990) *Science 249*:386-390); (Devlin (1990) *Science 249*:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci. 97*:6378-6382); (Felici (1991) *J. Mol. Biol. 222*:301-310); (Ladner *supra*).

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g.,

Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

Candidate compounds further include lysophospholipids, phospholipids, glycerophospholipids, sphingolipids, and lysosphingolipids. They can be related to natural ligands such as ceramide, sphingosine, S1P, LPA, cyclic LPA, cycosine, dihydrosphingosine, lysophosphatidyl-choline, lysophosphatidyl-ethanolamine, lysophosphatidyl serine, and lysosphingomyelin (sphingosyl-phosphorylcholine).

One candidate compound is a soluble full-length receptor or fragment that competes for ligand binding. Other candidate compounds include mutant receptors or appropriate fragments containing mutations that affect receptor function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention provides other end points to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the receptor protein, or a receptor protein target, could also be measured.

Targets in signaling include any of the intermediates in lipid-mediated GPCR transduction including adenyl cyclase, cAMP, receptor-G protein complex, G protein subunit disassociation, MAPK activation, activated Ras, P13Kγ, activated tyrosine kinases, Rho-activated Ser/Thr kinases, and phosphorylated MLC.

Any of the biological or biochemical functions mediated by the receptor can be used as an endpoint assay. These include all of the biochemicals or biochemical/biological events described herein, in the references cited herein,

incorporated by reference for these end point assay targets, and other functions known to those of ordinary skill in the art.

Binding and/or activating compounds can also be screened by using chimeric receptor proteins in which the amino terminal extracellular domain or part thereof, the region spanning the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops, and the carboxy terminal intracellular domain or part can be replaced by heterologous domains or parts thereof. For example, a G-protein-binding region can be used that interacts with a different G-protein then that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, one or more of the transmembrane segments or loops can be replaced with one or more of the transmembrane segments or loops specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the receptor is derived. Alternatively, the amino terminal extracellular domain or a part thereof and/or other ligand-binding regions could be replaced by a domain or part thereof and/or other ligand-binding regions binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The receptor polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the receptor. Thus, a compound is exposed to a receptor polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble receptor polypeptide is also added to the mixture. If the test compound interacts with the soluble receptor polypeptide, it decreases the amount of complex formed or activity from the receptor target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the receptor. Thus, the soluble polypeptide that competes with the target receptor region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is desirable to immobilize either the receptor protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/14275 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a receptor-binding protein and a candidate compound are incubated in the receptor protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the receptor protein target molecule, or which are reactive with receptor protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Modulators of receptor protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the receptor pathway, by treating cells that express the 14275 receptor protein, such as

in mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes. These methods of treatment include the steps of administering the modulators of protein activity in a pharmaceutical composition as described herein, to a subject in need of such treatment.

The polypeptides are thus useful for treating a receptor -associated disorder characterized by aberrant expression or activity of a receptor protein. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity of the protein. In another embodiment, the method involves administering a protein as therapy to compensate for reduced or aberrant expression or activity of the protein.

Stimulation of protein activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased protein activity is likely to have a beneficial effect. Likewise, inhibition of protein activity is desirable in situations in which the protein is abnormally upregulated and/or in which decreased protein activity is likely to have a beneficial effect. In one example of such a situation, a subject has a disorder characterized by aberrant development or cellular differentiation. In another example of such a situation, the subject has a proliferative disease (e.g., cancer) or a disorder characterized by an aberrant hematopoietic response. In another example of such a situation, it is desirable to achieve tissue regeneration in a subject (e.g., where a subject has undergone brain or spinal cord injury and it is desirable to regenerate neuronal tissue in a regulated manner).

In yet another aspect of the invention, the proteins of the invention can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell 72*:223-232; Madura *et al.* (1993) *J. Biol. Chem. 268*:12046-12054; Bartel *et al.* (1993) *Biotechniques 14*:920-924; Iwabuchi *et al.* (1993) *Oncogene 8*:1693-1696; and Brent WO 94/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their activity.

The receptor polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the receptor protein, especially in mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes. Accordingly, methods are provided for detecting the presence, or levels of, the receptor protein in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the receptor protein such that the interaction can be detected.

One agent for detecting receptor protein is an antibody capable of selectively binding to receptor protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The receptor protein also provides a target for diagnosing active disease, or predisposition to disease, in a patient having a variant receptor protein. Thus, receptor protein can be isolated from a biological sample, assayed for the presence of a genetic mutation that results in aberrant receptor protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein.

In vitro techniques for detection of receptor protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, the protein can be detected *in vivo* in a subject by introducing into the subject a labeled anti-receptor antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods which detect the allelic variant of a receptor protein expressed in a subject and methods which detect fragments of a receptor protein in a sample.

The receptor polypeptides are also useful in pharmacogenomic analysis. Accordingly, genetic polymorphism may lead to allelic protein variants of the receptor protein in which one or more of the receptor functions in one population is different from those in another population. The polypeptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic polypeptides could be identified.

The receptor polypeptides are also useful for monitoring therapeutic effects during clinical trials and other treatment. Thus, the therapeutic effectiveness of an agent that is designed to increase or decrease gene expression, protein levels or receptor activity can be monitored over the course of treatment using the receptor polypeptides as an end-point target.

The monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression or activity of a specified protein in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein in the post-administration samples; (v) comparing the level of expression or activity of the protein in the pre-administration sample with the protein in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The receptor polypeptides are also useful for treating a receptor-associated disorder. Accordingly, methods for treatment include the use of soluble receptor or fragments of the receptor protein that compete for ligand binding. These receptors or fragments can have a higher affinity for the ligand so as to provide effective competition.

Antibodies

The invention also provides antibodies that selectively bind to the 14275 receptor protein and its variants and fragments. An antibody is considered to

selectively bind, even if it also binds to other proteins that are not substantially homologous with the receptor protein. These other proteins share homology with a fragment or domain of the receptor protein. This conservation in specific regions gives rise to antibodies that bind to both proteins by virtue of the homologous sequence. In this case, it would be understood that antibody binding to the receptor protein is still selective.

Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g. Fab or $F(ab')_2$) can be used.

Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

To generate antibodies, an isolated receptor polypeptide is used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Figure 3 shows regions having a high antigenicity index. Antibodies are preferably prepared from these regions or from discrete fragments in these regions. However, antibodies can be prepared from any region of the peptide as described herein. A preferred fragment produces an antibody that diminishes or completely prevents ligand-binding. Antibodies can be developed against the entire receptor or portions of the receptor, for example, the intracellular carboxy terminal domain, the amino terminal extracellular domain, the entire transmembrane domain or specific segments, any of the intra or extracellular loops, or any portions of the above. Antibodies may also be developed against specific functional sites, such as the site of ligand-binding, the

site of G protein coupling, or sites that are phosphorylated, glycosylated, or myristoylated.

An antigenic fragment will typically comprise at least 6, 9, or 12 contiguous amino acid residues. The antigenic peptide can comprise a contiguous sequence of at least 14 amino acid residues, at least 15 amino acid residues, at least 20 amino acid residues, or at least 30 amino acid residues. In one embodiment, fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions. These fragments are not to be construed, however, as encompassing any fragments which may be disclosed prior to the invention.

An appropriate immunogenic preparation can be derived from native, recombinantly expressed, protein or chemically synthesized peptides.

Antibody Uses

The antibodies can be used to isolate a receptor protein by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural receptor protein from cells and recombinantly produced receptor protein expressed in host cells.

The antibodies are useful to detect the presence of receptor protein in cells or tissues to determine the pattern of expression of the receptor among various tissues in an organism and over the course of normal development.

The antibodies can be used to detect receptor protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression.

The antibodies can be used to assess abnormal tissue distribution or abnormal expression during development.

Antibody detection of circulating fragments of the full length receptor protein can be used to identify receptor turnover.

Further, the antibodies can be used to assess receptor expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to receptor function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, or level of expression of the receptor protein, the antibody can be prepared against the

normal receptor protein. If a disorder is characterized by a specific mutation in the receptor protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant receptor protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole receptor or portions of the receptor, for example, portions of the amino terminal extracellular domain or extracellular loops.

The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting receptor expression level or the presence of aberrant receptors and aberrant tissue distribution or developmental expression, antibodies directed against the receptor or relevant fragments can be used to monitor therapeutic efficacy.

Antibodies accordingly can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen.

Additionally, antibodies are useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol. 23 (10-11)*:983-985 (1996)); and Linder, M.W. (*Clin. Chem. 43(2)*:254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor

metabolizer. Thus, antibodies prepared against polymorphic receptor proteins can be used to identify individuals that require modified treatment modalities.

The antibodies are also useful as diagnostic tools as an immunological marker for aberrant receptor protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Thus, where a specific receptor protein has been correlated with expression in a specific tissue, antibodies that are specific for this receptor protein can be used to identify a tissue type.

The antibodies are also useful in forensic identification. Accordingly, where an individual has been correlated with a specific genetic polymorphism resulting in a specific polymorphic protein, an antibody specific for the polymorphic protein can be used as an aid in identification.

The antibodies are also useful for inhibiting receptor function, for example, blocking ligand binding.

These uses can also be applied in a therapeutic context in which treatment involves inhibiting receptor function. An antibody can be used, for example, to block ligand binding. Antibodies can be prepared against specific fragments containing sites required for function or against intact receptor associated with a cell.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol. 13*:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5.633,425; U.S. Patent 5,569.825; U.S. Patent 5,661,016; and U.S. Patent 5,545.806.

The invention also encompasses kits for using antibodies to detect the presence of a receptor protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting receptor protein in a biological sample; means for determining the amount of receptor protein in the sample; and means for comparing the amount of receptor protein in the sample with a standard. The compound or agent can be

packaged in a suitable container. The kit can further comprise instructions for using the kit to detect receptor protein.

Polynucleotides

The nucleotide sequence in SEQ ID NO 2 was obtained by sequencing the deposited human full length cDNA. Accordingly, the sequence of the deposited clone is controlling as to any discrepancies between the two and any reference to the sequence of SEQ ID NO 2 includes reference to the sequence of the deposited cDNA.

The specifically disclosed cDNA comprises the coding region, 5□ and 3□ untranslated sequences (SEQ ID NO 2). In one embodiment, the receptor nucleic acid comprises only the coding region.

The human 14275 receptor cDNA is approximately 1877 nucleotides in length and encodes a full length protein that is approximately 384 amino acid residues in length. The nucleic acid is expressed in: thymus, colon, spleen, and peripheral blood cells with lower expression in the lung, heart, small intestine, uterus, prostate, and placenta. Structural analysis of the amino acid sequence of SEQ ID NO 1 is provided in Figure 3, a hydropathy plot. The figure shows the putative structure of the seven transmembrane segments, the amino terminal extracellular domain and the carboxy terminal intracellular domain. As used herein, the term "transmembrane segment" refers to a structural amino acid motif which includes a hydrophobic helix that spans the plasma membrane. The entire transmembrane domain spans amino acids from about 51 to about 331. Seven segments span the membrane and there are three intracellular and three extracellular loops in the domain as explained for Figure 1.

The invention provides isolated polynucleotides encoding a 14275 receptor protein. The term "14275 polynucleotide" or "14275 nucleic acid" refers to the sequence shown in SEQ ID NO 2 or in the deposited cDNA. The term "receptor polynucleotide" or "receptor nucleic acid" further includes variants and fragments of the 14275 polynucleotide.

An "isolated" receptor nucleic acid is one that is separated from other nucleic acid present in the natural source of the receptor nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic

acid (i.e., sequences located at the 5′ and 3′ ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB. The important point is that the nucleic acid is isolated from flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the receptor nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA or RNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 % (on a molar basis) of all macromolecular species present.

The receptor polynucleotides can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

The receptor polynucleotides include, but are not limited to, the sequence encoding the mature polypeptide alone, the sequence encoding the mature polypeptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5′ and 3′ sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Receptor polynucleotides can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

One receptor nucleic acid comprises the nucleotide sequence shown in SEQ ID NO 2, corresponding to human prostate cDNA.

The invention further provides variant receptor polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO 2 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO 2.

The invention also provides receptor nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

Typically, variants have a substantial identity with a nucleic acid molecule of Figure 1 and the complements thereof.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. These variants comprise a nucleotide sequence encoding a receptor that is at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO 2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO 2 or a fragment of the sequence. It is understood that stringent hybridization does not indicate substantial homology where it is due to general homology, such as poly A sequences, or sequences common to all or most proteins, all GPCRs, all family I GPCRs, or all EDG receptors. Moreover, it is understood that variants do not include any of the nucleic acid sequences that may have been disclosed prior to the invention.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a receptor polypeptide at least 50-55%, 55% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 90%, at least about 95% or more identical to each other remain hybridized to one another. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, incorporated by reference. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. In another non-limiting example, nucleic acid molecules are allowed to hybridize in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more low stringency washes in 0.2X SSC/0.1% SDS at room temperature, or by one or more moderate stringency washes in 0.2X SSC/0.1%

SDS at 42°C, or washed in 0.2X SSC/0.1% SDS at 65°C for high stringency. In one embodiment, an isolated receptor nucleic acid molecule that hybridizes under stringent conditions to the sequence of Figure 1 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

As understood by those of ordinary skill, the exact conditions can be determined empirically and depend on ionic strength, temperature and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS. Other factors considered in determining the desired hybridization conditions include the length of the nucleic acid sequences, base composition, percent mismatch between the hybridizing sequences and the frequency of occurrence of subsets of the sequences within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules.

The present invention also provides isolated nucleic acids that contain a single or double stranded fragment or portion that hybridizes under stringent conditions to the nucleotide sequence of Figure 1 and the complements. In one embodiment, the nucleic acid consists of a portion of the nucleotide sequence of Figure 1 and the complements. The nucleic acid fragments of the invention are at least about 15, preferably at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, which encode antigenic proteins or polypeptides described herein are useful.

Furthermore, the invention provides polynucleotides that comprise a fragment of the full length receptor polynucleotides. The fragment can be single or double stranded and can comprise DNA or RNA. The fragment can be derived from either the coding or the non-coding sequence.

In one embodiment, an isolated receptor nucleic acid is at least 52 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO 2.

In another embodiment, an isolated receptor nucleic acid encodes the entire coding region from amino acid 1 to amino acid 384. In another

embodiment the isolated receptor nucleic acid encodes a sequence corresponding to the mature protein from about amino acid 6 to amino acid 384. Fragments further include nucleic acid sequences encoding a portion of the amino acid sequence described herein and further including flanking nucleotide sequences at the 3' region. Other fragments include nucleotide sequences encoding the amino acid fragments described herein. Further fragments can include subfragments of the specific domains or sites described herein. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof.

Receptor nucleic acid fragments also include a fragment from around nucleotide 1 to around 483 and subfragments thereof greater than 7 nucleotides. Receptor nucleic acid fragments further include a nucleotide sequence from around 477 to around 1143 and subfragments thereof greater than 18 nucleotides. A further receptor nucleic acid fragment includes nucleic acid from around 1121 to around 1369 and subfragments thereof greater than 33 nucleotides. A further fragment is from about 1387-1425 and subfragments thereof greater than 11 nucleotides. A further fragment is from about 1425 to the end of the sequence and subfragments thereof greater than 7 nucleotides. In these embodiments, the nucleic acid can be at least 17, 20, 30, 40, 50, 100, 250, or 500 nucleotides in length or greater. Nucleic acid fragments, according to the present invention, are not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

Receptor nucleic acid fragments further include sequences corresponding to the domains described herein, subregions also described, and specific functional sites. Receptor nucleic acid fragments include nucleic acid molecules encoding a polypeptide comprising the amino terminal extracellular domain including amino acid residues from 1 to about 50, a polypeptide comprising the region spanning the transmembrane domain (amino acid residues from about 51 to about 331), a polypeptide comprising the carboxy terminal intracellular domain (amino acid residues from about 332 to about 384), and a polypeptide encoding the G--protein receptor signature (142-144 or surrounding amino acid residues from about 135 to about 150), nucleic acid molecules encoding any of the seven transmembrane segments, extracellular or intracellular loops, glycosylation sites, protein kinase C

phosphorylation sites, and casein kinase II phosphorylation sites and myristoylation sites.

Receptor nucleic acid fragments also include combinations of the domains, segments, loops, and other functional sites described above. Thus, for example, a receptor nucleic acid could include sequences corresponding to the amino terminal extracellular domain and one transmembrane fragment. A person of ordinary skill in the art would be aware of the many permutations that are possible. Where the location of the domains have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

However, it is understood that a receptor fragment includes any nucleic acid sequence that does not include the entire gene.

The invention also provides receptor nucleic acid fragments that encode epitope bearing regions of the receptor proteins described herein.

The isolated receptor polynucleotide sequences, and especially fragments, are useful as DNA probes and primers.

For example, the coding region of a receptor gene can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of receptor genes.

A probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, typically about 25, more typically about 40, 50 or 75 consecutive nucleotides of SEQ ID NO 2 sense or anti-sense strand or other receptor polynucleotides. A probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

Polynucleotide Uses

The nucleic acid sequences of the present invention can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol. 215*:403-10. BLAST nucleotide searches can be

performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res. 25*(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The nucleic acid fragments of the invention provide probes or primers in assays such as those described below. "Probes" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid. Such probes include polypeptide nucleic acids, as described in Nielsen *et al.* (1991) *Science 254*:1497-1500. Typically, a probe comprises a region of nucleotide sequence that hybridizes under highly stringent conditions to at least about 15, typically about 20-25, and more typically about 40, 50 or 75 consecutive nucleotides of the nucleic acid sequence of SEQ ID NO 2 and the complements thereof. More typically, the probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis using well-known methods (e.g., PCR, LCR) including, but not limited to those described herein. The appropriate length of the primer depends on the particular use, but typically ranges from about 15 to 30 nucleotides. The term "primer site" refers to the area of the target DNA to which a primer hybridizes. The term "primer pair" refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the nucleic acid sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the sequence to be amplified.

The receptor polynucleotides are useful for probes, primers, and in biological assays.

Where the polynucleotides are used to assess GPCR properties or functions, such as in the assays described herein, all or less than all of the entire cDNA can be useful. In this case, even fragments that may have been known prior to the invention are encompassed. Thus, for example, assays specifically directed to GPCR functions, such as assessing agonist or antagonist activity,

encompass the use of known fragments. Further, diagnostic methods for assessing receptor function can also be practiced with any fragment, including those fragments that may have been known prior to the invention. Similarly, in methods involving treatment of receptor dysfunction, all fragments are encompassed including those which may have been known in the art.

The receptor polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptide described in SEQ ID NO 1 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptide shown in SEQ ID NO 1 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptide shown in SEQ ID NO 1 was isolated, different tissues from the same organism, or from different organisms. This method is useful for isolating genes and cDNA that are developmentally controlled and therefore may be expressed in the same tissue at different points in the development of an organism.

The probe can correspond to any sequence along the entire length of the gene encoding the receptor. Accordingly, it could be derived from 5′ noncoding regions, the coding region, and 3′ noncoding regions.

The nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO 1, or a fragment thereof, such as an oligonucleotide of at least 12, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA.

Fragments of the polynucleotides described herein are also useful to synthesize larger fragments or full-length polynucleotides described herein. For example, a fragment can be hybridized to any portion of an mRNA and a larger or full-length cDNA can be produced.

The fragments are also useful to synthesize antisense molecules of desired length and sequence.

Antisense nucleic acids of the invention can be designed using the nucleotide sequence of SEQ ID NO 2, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified

nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest.

Additionally, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry 4*:5). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996), *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA 93*:14670. PNAs can be further modified,

e.g., to enhance their stability, specificity or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra*, Finn *et al.* (1996) *Nucleic Acids Res. 24*(17):3357-63, Mag *et al.* (1989) *Nucleic Acids Res. 17*:5973, and Peterser *et al.* (1975) *Bioorganic Med. Chem. Lett. 5*:1119.

The nucleic acid molecules and fragments of the invention can also include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA 86*:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA 84*:648-652; PCT Publication No. WO 88/0918) or the blood brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques 6*:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm Res. 5*:539-549).

The receptor polynucleotides are also useful as primers for PCR to amplify any given region of a receptor polynucleotide.

The receptor polynucleotides are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the receptor polypeptides. Vectors also include insertion vectors, used to integrate into another polynucleotide sequence, such as into the cellular genome, to alter *in situ* expression of receptor genes and gene products. For example, an endogenous receptor coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The receptor polynucleotides are also useful as probes for determining the chromosomal positions of the receptor polynucleotides by means of *in situ* hybridization methods, such as FISH (for a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York), and PCR mapping of somatic cell hybrids. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature 325*:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible form chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

The receptor polynucleotide probes are also useful to determine patterns of the presence of the gene encoding the receptors and their variants with respect to tissue distribution, for example whether gene duplication has occurred and whether the duplication occurs in all or only a subset of tissues. The genes can be naturally occurring or can have been introduced into a cell, tissue, or organism exogenously. The receptor polynucleotides are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from genes encoding the polynucleotides described herein.

The receptor polynucleotides are also useful for constructing host cells expressing a part, or all, of the receptor polynucleotides and polypeptides.

The receptor polynucleotides are also useful for constructing transgenic animals expressing all, or a part, of the receptor polynucleotides and polypeptides.

The receptor polynucleotides are also useful for making vectors that express part, or all, of the receptor polypeptides.

The receptor polynucleotides are also useful as hybridization probes for determining the level of receptor nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, receptor nucleic acid in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant in cases in which there has been an amplification of the receptor genes.

Alternatively, the probe can be used in an *in situ* hybridization context to assess the position of extra copies of the receptor genes, as on extrachromosomal elements or as integrated into chromosomes in which the receptor gene is not normally found, for example as a homogeneously staining region.

These uses are relevant for diagnosis of disorders involving an increase or decrease in receptor expression relative to normal results, such as a proliferative disorder, a differentiative or developmental disorder, or a hematopoietic disorder, such as, for example, the disorders disclosed in the Example herein.

Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant expression or activity of receptor nucleic acid, in which a test sample is obtained from a subject and nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of the nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the nucleic acid.

One aspect of the invention relates to diagnostic assays for determining nucleic acid expression as well as activity in the context of a biological sample (e.g., blood, serum, cells, tissue) to determine whether an individual has a disease or disorder, or is at risk of developing a disease or disorder, associated with aberrant nucleic acid expression or activity. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual

prior to the onset of a disorder characterized by or associated with expression or activity of the nucleic acid molecules.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a receptor protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate receptor nucleic acid expression (e.g., antisense, polypeptides, peptidomimetics, small molecules or other drugs). A cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of receptor mRNA in the presence of the candidate compound is compared to the level of expression of receptor mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. The modulator can bind to the nucleic acid or indirectly modulate expression, such as by interacting with other cellular components that affect nucleic acid expression.

Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject) in patients or in transgenic animals.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the receptor gene. The method typically includes assaying the ability of the compound to modulate the expression of the receptor nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired receptor nucleic acid expression.

The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the receptor nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

Alternatively, candidate compounds can be assayed *in vivo* in patients or in transgenic animals.

The assay for receptor nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway (such as cyclic AMP or phosphatidylinositol turnover). Further, the expression of genes that are up- or down-regulated in response to the receptor protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase. Thus, modulators of receptor gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of receptor mRNA in the presence of the candidate compound is compared to the level of expression of receptor mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate receptor nucleic acid expression. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or effects on nucleic acid activity (e.g. when nucleic acid is mutated or improperly modified). Treatment is of disorders characterized by aberrant expression or activity of the nucleic acid.

Alternatively, a modulator for receptor nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the receptor nucleic acid expression.

The receptor polynucleotides are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the receptor gene in clinical trials or in a treatment regimen. Thus, the gene

expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

Monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a specified mRNA or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the mRNA or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the mRNA or genomic DNA in the pre-administration sample with the mRNA or genomic DNA in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The receptor polynucleotides are also useful in diagnostic assays for qualitative changes in receptor nucleic acid, and particularly in qualitative changes that lead to pathology. The polynucleotides can be used to detect mutations in receptor genes and gene expression products such as mRNA. The polynucleotides can be used as hybridization probes to detect naturally occurring genetic mutations in the receptor gene and thereby determining whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement such as inversion or transposition, modification of genomic DNA such as aberrant methylation patterns or changes in gene copy number such as amplification. Detection of a mutated form of the receptor gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a receptor protein.

Mutations in the receptor gene can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way.

In certain embodiments, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA 87*:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA 86*:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology 6*:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

Alternatively, mutations in a receptor gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S.Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.

Furthermore, sequence differences between a mutant receptor gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985); Cotton *et al.*, *Proc. Nat'l. Acad. Sci. USA 85*:4397 (1988); Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *Proc. Nat'l. Acad. Sci. USA 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature 313*:495 (1985)). The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet. 7*:5). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin *et al.* (1996) *Human Mutation* 7:244-255; Kozal *et al.* (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

The receptor polynucleotides are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the polynucleotides can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). In the present case, for example, a mutation in the receptor gene that results in altered affinity for ligand could result in an excessive or decreased drug effect with standard concentrations of ligand that activates the receptor. Accordingly, the receptor polynucleotides described herein can be used to assess the mutation content of the receptor gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus polynucleotides displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The methods can involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting mRNA, or genomic DNA, such that the presence of mRNA or

genomic DNA is detected in the biological sample, and comparing the presence of mRNA or genomic DNA in the control sample with the presence of mRNA or genomic DNA in the test sample.

The receptor polynucleotides are also useful for chromosome identification when the sequence is identified with an individual chromosome and to a particular location on the chromosome. First, the DNA sequence is matched to the chromosome by *in situ* or other chromosome-specific hybridization. Sequences can also be correlated to specific chromosomes by preparing PCR primers that can be used for PCR screening of somatic cell hybrids containing individual chromosomes from the desired species. Only hybrids containing the chromosome containing the gene homologous to the primer will yield an amplified fragment. Sublocalization can be achieved using chromosomal fragments. Other strategies include prescreening with labeled flow-sorted chromosomes and preselection by hybridization to chromosome-specific libraries. Further mapping strategies include fluorescence *in situ* hybridization which allows hybridization with probes shorter than those traditionally used. Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on the chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

The receptor polynucleotides can also be used to identify individuals from small biological samples. This can be done for example using restriction fragment-length polymorphism (RFLP) to identify an individual. Thus, the polynucleotides described herein are useful as DNA markers for RFLP (See U.S. Patent No. 5,272,057).

Furthermore, the receptor sequence can be used to provide an alternative technique which determines the actual DNA sequence of selected fragments in the genome of an individual. Thus, the receptor sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify DNA from an individual for subsequent sequencing.

Panels of corresponding DNA sequences from individuals prepared in this manner can provide unique individual identifications, as each individual will have a unique set of such DNA sequences. It is estimated that allelic variation in humans occurs with a frequency of about once per each 500 bases. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. The receptor sequences can be used to obtain such identification sequences from individuals and from tissue. The sequences represent unique fragments of the human genome. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

If a panel of reagents from the sequences is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

The receptor polynucleotides can also be used in forensic identification procedures. PCR technology can be used to amplify DNA sequences taken from very small biological samples, such as a single hair follicle, body fluids (e.g. blood, saliva, or semen). The amplified sequence can then be compared to a standard allowing identification of the origin of the sample.

The receptor polynucleotides can thus be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to the noncoding region are particularly useful since greater polymorphism occurs in the noncoding regions, making it easier to differentiate individuals using this technique. Fragments are at least 12 bases.

The receptor polynucleotides can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This is useful in cases in which a forensic pathologist is presented with a tissue of unknown origin. Panels

of receptor probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these primers and probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Alternatively, the receptor polynucleotides can be used directly to block transcription or translation of receptor gene expression by means of antisense or ribozyme constructs. Thus, in a disorder characterized by abnormally high or undesirable receptor gene expression, nucleic acids can be directly used for treatment.

The receptor polynucleotides are thus useful as antisense constructs to control receptor gene expression in cells, tissues, and organisms. A DNA antisense polynucleotide is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of receptor protein. An antisense RNA or DNA polynucleotide would hybridize to the mRNA and thus block translation of mRNA into receptor protein.

Examples of antisense molecules useful to inhibit nucleic acid expression include antisense molecules complementary to a fragment of the 5' untranslated region of SEQ ID NO 2 which also includes the start codon and antisense molecules which are complementary to a fragment of the 3' untranslated region of SEQ ID NO 2.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of receptor nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired receptor nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the receptor protein.

The receptor polynucleotides also provide vectors for gene therapy in patients containing cells that are aberrant in receptor gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired receptor protein to treat the individual.

The invention also encompasses kits for detecting the presence of a receptor nucleic acid in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting receptor nucleic acid in a biological sample; means for determining the amount of receptor nucleic acid in the sample; and means for comparing the amount of receptor nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect receptor mRNA or DNA.

Computer Readable Means

The nucleotide or amino acid sequences of the invention are also provided in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a nucleotide or amino acid sequence of the present invention. Such a manufacture provides the nucleotide or amino acid sequences, or a subset thereof (e.g., a subset of open reading frames (ORFs)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. The skilled artisan can readily adopt any of the presently known methods for recording information on computer readable

medium to generate manufactures comprising the nucleotide or amino acid sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of dataprocessor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the

folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA).

For example, software which implements the BLAST (Altschul *et al.* (1990) *J. Mol. Biol. 215*:403-410) and BLAZE (Brutlag *et al.* (1993) *Comp. Chem. 17*:203-207) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzymes used in various reactions and in the production of commercially useful metabolites.

Vectors/host cells

The invention also provides vectors containing the receptor polynucleotides. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, that can transport the receptor polynucleotides. When the vector is a nucleic acid molecule, the receptor polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the receptor polynucleotides. Alternatively, the vector may integrate into the host cell genome

and produce additional copies of the receptor polynucleotides when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the receptor polynucleotides. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the receptor polynucleotides such that transcription of the polynucleotides is allowed in a host cell. The polynucleotides can be introduced into the host cell with a separate polynucleotide capable of affecting transcription. Thus, the second polynucleotide may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the receptor polynucleotides from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself.

It is understood, however, that in some embodiments, transcription and/or translation of the receptor polynucleotides can occur in a cell-free system.

The regulatory sequence to which the polynucleotides described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in

Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a receptor polynucleotide. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The receptor polynucleotides can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the polypeptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for

the production of the receptor polypeptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired polypeptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.* (1988) *Gene 67*:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene 69*:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology 185*:119-128, Academic Press, San Diego, California (1990)). Alternatively, the sequence of the polynucleotide of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res. 20*:2111-2118 (1992)).

The receptor polynucleotides can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943(1982)), pJRY88 (Schultz et al., Gene 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The receptor polynucleotides can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the polynucleotides described herein are expressed in mammalian cells using mammalian expression vectors.

Examples of mammalian expression vectors include pCDM8 (Seed, B., *Nature 329*:840 (1987)) and pMT2PC (Kaufman *et al.*, *EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the receptor polynucleotides. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example in Sambrook, J.; Fritsh, E. F.; and Maniatis, T., *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the receptor polynucleotides can be introduced either alone or with other polynucleotides that are not related to the receptor polynucleotides such as those

providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the receptor polynucleotide vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the polypeptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the receptor polypeptides or heterologous to these polypeptides.

Where the polypeptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The polypeptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the polypeptides described herein, the polypeptides can have various glycosylation patterns, depending upon the cell, or maybe non-

glycosylated as when produced in bacteria. In addition, the polypeptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

It is understood that "host cells" and "recombinant host cells" refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

The host cells expressing the polypeptides described herein, and particularly recombinant host cells, have a variety of uses. First, the cells are useful for producing receptor proteins or polypeptides that can be further purified to produce desired amounts of receptor protein or fragments. Thus, host cells containing expression vectors are useful for polypeptide production.

Host cells are also useful for conducting cell-based assays involving the receptor or receptor fragments. Thus, a recombinant host cell expressing a native receptor is useful to assay for compounds that stimulate or inhibit receptor function. This includes ligand binding, gene expression at the level of transcription or translation, G-protein interaction, and components of the signal transduction pathway.

Cell-based assays include NE-115 (Postma, cited above); *Xenopus* oocytes, especially for calcium efflux (An, *FEBS Lett.*, cited above) and Cl currents (Guo, cited above); Jurkat cells, especially for reporter assays using SRE-driven transcription (An, *FEBS Lett.*, cited above); HEK 293 and CHO cells, especially for reporter assays using SRE-driven transcription (An, *Biochem. Biophys. Res. Comm.*, cited above).

Host cells are also useful for identifying receptor mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant receptor (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native receptor.

Recombinant host cells are also useful for expressing the chimeric polypeptides described herein to assess compounds that activate or suppress activation by means of a heterologous amino terminal extracellular domain (or other binding region). Alternatively, a heterologous region spanning the entire transmembrane domain (or parts thereof) can be used to assess the effect of a desired amino terminal extracellular domain (or other binding region) on any given host cell. In this embodiment, a region spanning the entire transmembrane domain (or parts thereof) compatible with the specific host cell is used to make the chimeric vector. Alternatively, a heterologous carboxy terminal intracellular, e.g., signal transduction, domain can be introduced into the host cell.

Further, mutant receptors can be designed in which one or more of the various functions is engineered to be increased or decreased (i.e., ligand binding or G-protein binding) and used to augment or replace receptor proteins in an individual. Thus, host cells can provide a therapeutic benefit by replacing an aberrant receptor or providing an aberrant receptor that provides a therapeutic result. In one embodiment, the cells provide receptors that are abnormally active.

In another embodiment, the cells provide receptors that are abnormally inactive. These receptors can compete with endogenous receptors in the individual.

In another embodiment, cells expressing receptors that cannot be activated, are introduced into an individual in order to compete with endogenous receptors for ligand. For example, in the case in which excessive ligand is part of a treatment modality, it may be necessary to inactivate this ligand at a specific point in treatment. Providing cells that compete for the ligand, but which cannot be affected by receptor activation would be beneficial.

Homologously recombinant host cells can also be produced that allow the *in situ* alteration of endogenous receptor polynucleotide sequences in a host cell genome. The host cell includes, but is not limited to, a stable cell line, cell *in vivo*, or cloned microorganism. This technology is more fully described in WO 93/09222, WO 91/12650, WO 91/06667, U.S. 5,272,071, and U.S. 5,641,670. Briefly, specific polynucleotide sequences corresponding to the receptor polynucleotides or sequences proximal or distal to a receptor gene are allowed to integrate into a host cell genome by homologous recombination where expression of the gene can be affected. In one embodiment, regulatory sequences are introduced that either

increase or decrease expression of an endogenous sequence. Accordingly, a receptor protein can be produced in a cell not normally producing it. Alternatively, increased expression of receptor protein can be effected in a cell normally producing the protein at a specific level. Further, expression can be decreased or eliminated by introducing a specific regulatory sequence. The regulatory sequence can be heterologous to the receptor protein sequence or can be a homologous sequence with a desired mutation that affects expression. Alternatively, the entire gene can be deleted. The regulatory sequence can be specific to the host cell or capable of functioning in more than one cell type. Still further, specific mutations can be introduced into any desired region of the gene to produce mutant receptor proteins. Such mutations could be introduced, for example, into the specific functional regions such as the ligand-binding site.

In one embodiment, the host cell can be a fertilized oocyte or embryonic stem cell that can be used to produce a transgenic animal containing the altered receptor gene. Alternatively, the host cell can be a stem cell or other early tissue precursor that gives rise to a specific subset of cells and can be used to produce transgenic tissues in an animal. See also Thomas *et al.*, *Cell 51*:503 (1987) for a description of homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g.. by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous receptor gene is selected (see e.g., Li, E. *et al.*, *Cell 69*:915 (1992)). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson. ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A., *Current Opinions in Biotechnology 2*:823-829 (1991); and in PCT International Publication Nos. WO 90/11354; WO 91/01140; and WO 93/04169.

The genetically engineered host cells can be used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a

rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a receptor protein and identifying and evaluating modulators of receptor protein activity.

Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which receptor polynucleotide sequences have been introduced.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the receptor nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the receptor protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986)). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the

entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* (*Proc. Nat'l. Acad. Sci. USA 89*:6232-6236 (1992)). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al.*, *Science 251*:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (*Nature 385*:810-813 (1997)); and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the polypeptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, receptor activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* receptor function, including ligand interaction, the effect of specific mutant receptors on receptor function and ligand interaction, and the effect of chimeric

receptors. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more receptor functions.

In general, methods for producing transgenic animals include introducing a nucleic acid sequence according to the present invention, the nucleic acid sequence capable of expressing the receptor protein in a transgenic animal, into a cell in culture or *in vivo*. When introduced *in vivo*, the nucleic acid is introduced into an intact organism such that one or more cell types and, accordingly, one or more tissue types, express the nucleic acid encoding the receptor protein. Alternatively, the nucleic acid can be introduced into virtually all cells in an organism by transfecting a cell in culture, such as an embryonic stem cell, as described herein for the production of transgenic animals, and this cell can be used to produce an entire transgenic organism. As described, in a further embodiment, the host cell can be a fertilized oocyte. Such cells are then allowed to develop in a female foster animal to produce the transgenic organism.

Pharmaceutical compositions

The receptor nucleic acid molecules, protein (particularly fragments such as the amino terminal extracellular domain), modulators of the protein, and antibodies (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier.

As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the compositions of the invention. Supplementary active compounds can also be incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal

(topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a receptor protein or anti-receptor antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For oral administration, the agent can be contained in enteric forms to survive the stomach or further coated or mixed to be released in a particular region of the GI tract by known methods. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are

generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. 5,328,470) or by stereotactic injection (see e.g., Chen *et al.*, *PNAS 91*:3054-3057

(1994)). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less

than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

This invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will fully convey the invention to those skilled in the art. Many modifications and other embodiments of the invention will come to mind in one skilled in the art to which this invention

pertains having the benefit of the teachings presented in the foregoing description. Although specific terms are employed, they are used as in the art unless otherwise indicated.

EXAMPLE

The expression of the 14275 receptor was studied using a Taqman analysis. This procedure involves RT-PCR, according to routine procedures. Extremely high expression was observed in mobilized peripheral blood $CD34^+$ cells. High expression was observed in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, and leukocytes. Significant expression was also observed in spleen, $CD34^-$ mobilized peripheral blood cells, $CD4^+$T lymphocytes, resting B lymphocytes, and resting peripheral blood mononuclear cells. There was higher expression in resting peripheral blood mononuclear cells than in PHA-activated periperhal blood mononuclear cells. Some expression was also observed in granulocytes. In T lymphocytes (Th1 and Th2) stimulated with anti-CD3, expression decreased over time (6-48 hours).

In view of these expression data, expression of the 14275 receptor is relevant to immunological disorders and disorders involving inflammation. Such immune disorders include, but are not limited to, chronic inflammatory diseases and disorders, such as Crohn's disease, reactive arthritis, including Lyme disease, insulin-dependent diabetes, organ-specific autoimmunity, including multiple sclerosis, Hashimoto's thyroiditis and Grave's disease, contact dermatitis, psoriasis, graft rejection, graft versus host disease, sarcoidosis, atopic conditions, such as asthma and allergy, including allergic rhinitis, gastrointestinal allergies, including food allergies, eosinophilia, conjunctivitis, glomerular nephritis, certain pathogen susceptibilities such as helminthic (e.g., leishmaniasis), certain viral infections, including HIV, and bacterial infections, including tuberculosis and lepromatous leprosy.

Respiratory disorders include, but are not limited to, apnea, asthma, particularly bronchial asthma, berillium disease, bronchiectasis, bronchitis, bronchopneumonia, cystic fibrosis, diphtheria, dyspnea, emphysema, chronic obstructive pulmonary disease, allergic bronchopulmonary aspergillosis,

pneumonia, acute pulmonary edema, pertussis, pharyngitis, atelectasis, Wegener's granulomatosis, Legionnaires disease, pleurisy, rheumatic fever, and sinusitis.

Further, in view of expression of the receptor in blood progenitor cells, expression of the receptor is relevant to blood cell formation and thus useful for treating and diagnosing anemia, neutropenia, and thrombocytopenia.

Analysis of expression was also done using standard northern blotting procedures. Expression was observed in peripheral blood cells, such as T and B cells, spleen, lung, thymus, uterus, small intestine, colon, heart, prostate and placenta.

Accordingly, the methods disclosed herein, including but not limited to, methods for identifying agents that modulate the level or activity of the receptor nucleic acid or polypeptide in a cell, methods of screening a cell to identify an agent that modulates the level or activity of the polypeptide or nucleic acid in a cell, methods for identifying agents that interact with the polypeptide or nucleic acid in a cell, methods of screening a cell to identify an agent that interacts with the polypeptide or nucleic acid in a cell, methods for detecting the presence of the polypeptide or nucleic acid in a cell, methods for modulating the level or activity of the polypeptide or nucleic acid in a cell, and any method of diagnosis and treatment based on these generic methods, are particularly applicable for these disorders and for and in the cell types in which the receptor is expressed or in which expression is abnormally low or absent.

In normal bone marrow, the myelocytic series (polymorphoneuclear cells) make up approximately 60% of the cellular elements, and the erythrocytic series, 20-30%. Lymphocytes, monocytes, reticular cells, plasma cells and megakaryocytes together constitute 10-20%. Lymphocytes make up 5-15% of normal adult marrow. In the bone marrow, cell types are add mixed so that precursors of red blood cells (erythroblasts), macrophages (monoblasts), platelets (megakaryocytes), polymorphoneuclear leucocytes (myeloblasts), and lymphocytes (lymphoblasts) can be visible in one microscopic field. In addition, stem cells exist for the different cell lineages, as well as a precursor stem cell for the committed progenitor cells of the different lineages. The various types of cells and stages of each would be known to the person of ordinary skill in the art and are found, for example, on page 42 (Figure 2-8) of *Immunology, Imunopathology and Immunity*, Fifth Edition, Sell *et al*. Simon and Schuster (1996), incorporated

by reference for its teaching of cell types found in the bone marrow. According, the invention is directed to disorders arising from these cells. These disorders include but are not limited to the following: diseases involving hematopoetic stem cells; committed lymphoid progenitor cells; lymphoid cells including B and T-cells; committed myeloid progenitors, including monocytes, granulocytes, and megakaryocytes; and committed erythroid progenitors. These include but are not limited to the leukemias, including B-lymphoid leukemias, T-lymphoid leukemias, undifferentiated leukemias; erythroleukemia, megakaryoblastic leukemia, monocytic; [leukemias are encompassed with and without differentiation]; chronic and acute lymphoblastic leukemia, chronic and acute lymphocytic leukemia, chronic and acute myelogenous leukemia, lymphoma, myelo dysplastic syndrome, chronic and acute myeloid leukemia, myelomonocytic leukemia; chronic and acute myeloblastic leukemia, chronic and acute myelogenous leukemia, chronic and acute promyelocytic leukemia, chronic and acute myelocytic leukemia, hematologic malignancies of monocyte-macrophage lineage, such as juvenile chronic myelogenous leukemia; secondary AML, antecedent hematological disorder; refractory anemia; aplastic anemia; reactive cutaneous angioendotheliomatosis; fibrosing disorders involving altered expression in dendritic cells, disorders including systemic sclerosis, E-M syndrome, epidemic toxic oil syndrome, eosinophilic fasciitis localized forms of scleroderma, keloid, and fibrosing colonopathy; angiomatoid malignant fibrous histiocytoma; carcinoma, including primary head and neck squamous cell carcinoma; sarcoma, including kaposi's sarcoma; fibroadanoma and phyllodes tumors, including mammary fibroadenoma; stromal tumors; phyllodes tumors, including histiocytoma; erythroblastosis; neurofibromatosis; diseases of the vascular endothelium; demyelinating, particularly in old lesions; gliosis, vasogenic edema, vascular disease, Alzheimer's and Parkinson's disease; T-cell lymphomas; B-cell lymphomas.

Disorders involving T-cells include, but are not limited to, cell-mediated hypersensitivity, such as delayed type hypersensitivity and T-cell-mediated cytotoxicity, and transplant rejection; autoimmune diseases, such as systemic lupus erythematosus, Sjögren syndrome, systemic sclerosis, inflammatory myopathies, mixed connective tissue disease, and polyarteritis nodosa and other vasculitides; immunologic deficiency syndromes, including but not limited to,

primary immunodeficiencies, such as thymic hypoplasia, severe combined immunodeficiency diseases, and AIDS; leukopenia; reactive (inflammatory) proliferations of white cells, including but not limited to, leukocytosis, acute nonspecific lymphadenitis, and chronic nonspecific lymphadenitis; neoplastic proliferations of white cells, including but not limited to lymphoid neoplasms, such as precursor T-cell neoplasms, such as acute lymphoblastic leukemia/lymphoma, peripheral T-cell and natural killer cell neoplasms that include peripheral T-cell lymphoma, unspecified, adult T-cell leukemia/lymphoma, mycosis fungoides and Sézary syndrome, and Hodgkin disease.

Disorders involving red cells include, but are not limited to, anemias, such as hemolytic anemias, including hereditary spherocytosis, hemolytic disease due to erythrocyte enzyme defects: glucose-6-phosphate dehydrogenase deficiency, sickle cell disease, thalassemia syndromes, paroxysmal nocturnal hemoglobinuria, immunohemolytic anemia, and hemolytic anemia resulting from trauma to red cells; and anemias of diminished erythropoiesis, including megaloblastic anemias, such as anemias of vitamin B12 deficiency: pernicious anemia, and anemia of folate deficiency, iron deficiency anemia, anemia of chronic disease, aplastic anemia, pure red cell aplasia, and other forms of marrow failure.

Disorders involving B-cells include, but are not limited to precursor B-cell neoplasms, such as lymphoblastic leukemia/lymphoma. Peripheral B-cell neoplasms include, but are not limited to, chronic lymphocytic leukemia/small lymphocytic lymphoma, follicular lymphoma, diffuse large B-cell lymphoma, Burkitt lymphoma, plasma cell neoplasms, multiple myeloma, and related entities, lymphoplasmacytic lymphoma (Waldenstrōm macroglobulinemia), mantle cell lymphoma, marginal zone lymphoma (MALToma), and hairy cell leukemia.

Disorders related to reduced platelet number (thrombocytopenia) include idiopathic thrombocytopenic purpura, including acute idiopathic thrombocytopenic purpura, drug-induced thrombocytopenia, HIV-associated thrombocytopenia, and thrombotic microangiopathies: thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome.

Disorders involving precursor T-cell neoplasms include precursor T lymphoblastic leukemia/lymphoma. Disorders involving peripheral T-cell and natural killer cell neoplasms include T-cell chronic lymphocytic leukemia, large

granular lymphocytic leukemia, mycosis fungoides and Sézary syndrome, peripheral T-cell lymphoma, unspecified, angioimmunoblastic T-cell lymphoma, angiocentric lymphoma (NK/T-cell lymphoma[4a]), intestinal T-cell lymphoma, adult T-cell leukemia/lymphoma, and anaplastic large cell lymphoma.

Disorders involving the spleen include, but are not limited to, splenomegaly, including nonspecific acute splenitis, congestive spenomegaly, and spenic infarcts; neoplasms, congenital anomalies, and rupture. Disorders associated with splenomegaly include infections, such as nonspecific splenitis, infectious mononucleosis, tuberculosis, typhoid fever, brucellosis, cytomegalovirus, syphilis, malaria, histoplasmosis, toxoplasmosis, kala-azar, trypanosomiasis, schistosomiasis, leishmaniasis, and echinococcosis; congestive states related to partial hypertension, such as cirrhosis of the liver, portal or splenic vein thrombosis, and cardiac failure; lymphohematogenous disorders, such as Hodgkin disease, non-Hodgkin lymphomas/leukemia, multiple myeloma, myeloproliferative disorders, hemolytic anemias, and thrombocytopenic purpura; immunologic-inflammatory conditions, such as rheumatoid arthritis and systemic lupus erythematosus; storage diseases such as Gaucher disease, Niemann-Pick disease, and mucopolysaccharidoses; and other conditions, such as amyloidosis, primary neoplasms and cysts, and secondary neoplasms.

Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis; diffuse interstitial (infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia), *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary hemosiderosis and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and

lung transplantation; tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

Disorders involving the colon include, but are not limited to, congenital anomalies, such as atresia and stenosis, Meckel diverticulum, congenital aganglionic megacolon-Hirschsprung disease; enterocolitis, such as diarrhea and dysentery, infectious enterocolitis, including viral gastroenteritis, bacterial enterocolitis, necrotizing enterocolitis, antibiotic-associated colitis (pseudomembranous colitis), and collagenous and lymphocytic colitis, miscellaneous intestinal inflammatory disorders, including parasites and protozoa, acquired immunodeficiency syndrome, transplantation, drug-induced intestinal injury, radiation enterocolitis, neutropenic colitis (typhlitis), and diversion colitis; idiopathic inflammatory bowel disease, such as Crohn disease and ulcerative colitis; tumors of the colon, such as non-neoplastic polyps, adenomas, familial syndromes, colorectal carcinogenesis, colorectal carcinoma, and carcinoid tumors.

Disorders involving the uterus and endometrium include, but are not limited to, endometrial histology in the menstrual cycle; functional endometrial disorders, such as anovulatory cycle, inadequate luteal phase, oral contraceptives and induced endometrial changes, and menopausal and postmenopausal changes; inflammations, such as chronic endometritis; adenomyosis; endometriosis; endometrial polyps; endometrial hyperplasia; malignant tumors, such as carcinoma of the endometrium; mixed Müllerian and mesenchymal tumors, such as malignant mixed Müllerian tumors; tumors of the myometrium, including leiomyomas, leiomyosarcomas, and endometrial stromal tumors.

Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular

degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

Disorders involving the thymus include developmental disorders, such as DiGeorge syndrome with thymic hypoplasia or aplasia; thymic cysts; thymic hypoplasia, which involves the appearance of lymphoid follicles within the thymus, creating thymic follicular hyperplasia; and thymomas, including germ cell tumors, lynphomas, Hodgkin disease, and carcinoids. Thymomas can include benign or encapsulated thymoma, and malignant thymoma Type I (invasive thymoma) or Type II, designated thymic carcinoma.

Disorders involving the prostate include, but are not limited to, inflammations, benign enlargement, for example, nodular hyperplasia (benign prostatic hypertrophy or hyperplasia), and tumors such as carcinoma.

THAT WHICH IS CLAIMED:

1. An isolated polypeptide having an amino acid sequence selected from the group consisting of:

(a) The amino acid sequence shown in SEQ ID NO 1;

(b) The amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____;

(c) The amino acid sequence of an allelic variant of the amino acid sequence shown in SEQ ID NO 1;

(d) The amino acid sequence of an allelic variant of the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____;

(e) The amino acid sequence of a sequence variant of the amino acid sequence shown in SEQ ID NO 1, wherein the sequence variant is encoded by a nucleic acid molecule hybridizing to the nucleic acid molecule shown in SEQ ID NO 2 under stringent conditions;

(f) The amino acid sequence of a sequence variant of the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No.___, wherein the sequence variant is encoded by a nucleic acid molecule hybridizing under stringent conditions to the cDNA contained in ATCC Deposit No.____;

(g) A fragment of the amino acid sequence shown in SEQ ID NO 1, wherein the fragment comprises at least 18 contiguous amino acids;

(h) A fragment of the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____, wherein the fragment comprises at least 18 contiguous amino acids;

(i) The amino acid sequence of the mature receptor polypeptide from about amino acid 6 to about amino acid 384, shown in SEQ ID NO 1;

(j) The amino acid sequence of the mature polypeptide from about amino acid 6 to about amino acid 384, encoded by the cDNA clone contained in ATCC Deposit No. _____;

(k) The amino acid sequence of the polypeptide shown in SEQ ID NO 1, from about amino acid 332 to about amino acid 384;

(l) The amino acid sequence from about amino acid 332 to about amino acid 384 in the polypeptide encoded by the cDNA contained in ATCC Deposit No. _____;

(m) The amino acid sequence of an epitope bearing region of any one of the polypeptides of (a)-(l).

2. An isolated antibody that selectively binds to a polypeptide of claim 1, (a)-(m).

3. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:

(a) The nucleotide sequence shown in SEQ ID NO 2;

(b) The nucleotide sequence in the cDNA contained in ATCC Deposit No. ____;

(c) A nucleotide sequence encoding the amino acid sequence shown in SEQ ID NO 1;

(d) A nucleotide sequence encoding the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____; and

(e) A nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), or (d).

4. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:

(a) A nucleotide sequence encoding an amino acid sequence of a sequence variant of the amino acid sequence shown in SEQ ID NO 1 that hybridizes to the nucleotide sequence shown in SEQ ID NO 2 under stringent conditions;

(b) A nucleotide sequence encoding the amino acid sequence of a sequence variant of the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____, the nucleic acid sequence of the sequence variant hybridizing to the cDNA contained in ATCC Deposit No. _____ under stringent conditions; and

(c) A nucleotide sequence complementary to either of the nucleotide sequences in (a) or (b).

5. An isolated nucleic acid molecule a polynucleotide having a nucleotide sequence selected from the group consisting of:

(a) A nucleotide sequence encoding a fragment of the amino acid sequence shown in SEQ ID NO 1, wherein the fragment comprises at least 18 contiguous amino acids;

(b) A nucleotide sequence encoding a fragment of the amino acid sequence encoded by the cDNA contained in ATCC Deposit No. _____, wherein the fragment comprises at least 18 contiguous amino acids;

(c) A nucleotide sequence complementary to either of the nucleotide sequences in (a) or (b).

6. A nucleic acid vector comprising the nucleic acid sequences in any of claims 3-5.

7. A host cell containing the vector of claim 6.

8. A method for producing any of the polypeptides in claim 1 comprising introducing a nucleotide sequence encoding any of the polypeptide sequences in (a)-(m) into a host cell, and culturing the host cell under conditions in which the proteins are expressed from the nucleic acid.

9.. A method for identifying an agent that binds to any of the polypeptides in claim 1, said method comprising contacting the polypeptide with an agent that binds to the polypeptide and assaying the complex formed with the agent bound to the polypeptide.

10. The method of claim 9, wherein a fragment of the polypeptide is contacted.

11. A method for identifying an agent that modulates the level or activity of any of the polypeptides of claim 1 in a cell, said method comprising: contacting said agent with a cell capable of expressing said polypeptide such that

said polypeptide level or activity can be modulated in said cell by said agent and measuring said polypeptide level or activity, wherein said cell is derived from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

12. A method of screening a cell to identify an agent that modulates the level or activity of any of the polypeptides of claim 1 in a cell, said method comprising: contacting said agent with a cell capable of expressing said polypeptide such that said polypeptide level or activity can be modulated in said cell by said agent and measuring said polypeptide level or activity, wherein said cell is derived from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

13. The method of claim 11 wherein said cell is a $CD34^+$ bone marrow cell.

14. The method of claim 11 wherein said agent increases the level or activity of said polypeptide.

15. The method of claim 11 wherein said agent decreases the level or activity of said polypeptide.

16. A method for identifying an agent that interacts with any of the polypeptides of claim 1 in a cell, said method comprising: contacting said agent with a cell capable of allowing an interaction between said polypeptide and said agent such that said polypeptide can interact with said agent and measuring the interaction, wherein said cell is derived from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$

mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

17. A method of screening a cell to identify an agent that interacts with any of the polypeptides of claim 1 in a cell, said method comprising: contacting said agent with a cell capable of allowing an interaction between said polypeptide and said agent such that said polypeptide can interact with said agent and measuring the interaction, wherein said cell is derived from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

18. The method of claim 16, said method comprising: (1) exposing said agent to said polypeptide under conditions that allow said agent to interact with said polypeptide; (2) adding competing polypeptide that can interact with said agent; and (3) comparing the amount of interaction between said agent and said polypeptide to the amount of interaction in the absence of said competing polypeptide.

19. The method of claim 16 wherein said interaction is binding.

20. The method of claim 11 wherein said agent increases interaction between said polypeptide and a target molecule for said polypeptide, said method comprising: combining said polypeptide with said agent under conditions that allow said polypeptide to interact with said target molecule; and detecting the formation of a complex between said polypeptide and said target molecule or activity of said polypeptide as a result of interaction of said polypeptide with said target molecule.

21. The method of claim 11 wherein said agent decreases interaction between said polypeptide and a target molecule for said polypeptide, said method comprising: combining said polypeptide with said agent under conditions that

allow said polypeptide to interact with said target molecule; and detecting the formation of a complex between said polypeptide and said target molecule or activity of said polypeptide as a result of interaction of said polypeptide with said target molecule.

22. The method of claim 11 wherein said cell is *in vivo*.

23. The method of claim 22 wherein said cell is in a transgenic animal.

24. The method of claim 22 wherein said cell is in a non-transgenic subject.

25. The method of claim 11 wherein said cell is *in vitro*.

26. The method of claim 25 wherein said cell has been disrupted.

27. The method of claim 25 wherein said cell is in a biopsy.

28. The method of claim 26 wherein said cell is in cell culture.

29. The method of claim 28 wherein said cell is naturally-occurring or recombinant.

30. The method of claim 11 wherein said agent is selected from the group consisting of a peptide; phosphopeptide; antibody; organic molecule; and inorganic molecule.

31. A method for detecting the presence of any of the polypeptides of claim 1 in a sample, said method comprising contacting said sample with an agent that specifically allows detection of the presence of the polypeptide in the sample and then detecting the presence of the polypeptide, wherein said sample is derived from a cell selected from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node,

leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

32. The method of claim 21, wherein said agent is capable of selective physical association with said polypeptide.

33. The method of claim 22, wherein said agent binds to said polypeptide.

34. The method of claim 23, wherein said agent is an antibody.

35. The method of claim 23, wherein said agent is a ligand.

36. A kit comprising reagents used for the method of claim 21, wherein the reagents comprise an agent that specifically binds to said polypeptide.

37. A method for modulating the level or activity of any of the polypeptides of claim 1, wherein said modulation occurs in cells derived from tissue selected from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

38. A method for identifying an agent that modulates the level or activity of any of the nucleic acid molecules of claims 3-5 in a cell, said method comprising contacting said agent with a cell capable of expressing said nucleic acid molecule such that said nucleic acid molecule level or activity can be modulated in said cell by said agent and measuring said nucleic acid molecule level or activity, wherein said cell is derived from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

39. A method of screening a cell to identify an agent that modulates the level or activity of any of the nucleic acid molecules of claims 3-5 in said cell, said method comprising: contacting said agent with a cell capable of expressing said nucleic acid molecule such that said nucleic acid molecule level or activity can be modulated in said cell by said agent and measuring nucleic acid molecule level or activity, wherein said cell is derived from the group consisting of brain, $CD34^+$, B, skeletal muscle, lymph node, spleen, thymus, liver, tonsils, colon, heart, granulocyte, erythroblast, and pancreatic cells.

40. A method for identifying an agent that interacts with any of the nucleic acid molecules of claims 3-5 in a cell, said method comprising: contacting said agent with a cell capable of allowing an interaction between said nucleic acid molecule and said agent such that said nucleic acid molecule can interact with said agent and measuring the interaction, wherein said cell is derived from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

41. A method of screening a cell to identify an agent that interacts with any of the nucleic acid molecules of claims 3-5 in a cell, said method comprising: contacting said agent with a cell capable of allowing an interaction between said nucleic acid molecule and said agent, such that nucleic acid molecule can interact with said agent and measuring the interaction, wherein said cell is derived from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

42. A method for detecting the presence of any of the nucleic acid molecules of claims 3-5 in a sample, said method comprising contacting said

sample with an agent that specifically allows detection of the presence of the nucleic acid molecule in the sample and then detecting the presence of the nucleic acid molecule, wherein said sample is derived from a tissue selected from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

43. The method of claim 42, wherein the method comprises contacting the sample with an oligonucleotide that hybridizes to any of the nucleic acid sequences of (a)-(k) under stringent conditions and determining whether the oligonucleotide binds to the nucleic acid sequence in the sample.

44. The method of claim 43, wherein the nucleic acid whose presence is detected is mRNA.

45. A kit comprising reagents used for the method of claim 43, wherein the reagents comprise a compound that hybridizes under stringent conditions to any of the nucleic acid molecules.

46. The method of claim 45 wherein a fragment of the nucleic acid is contacted.

47. A method for modulating the level or activity of any of the nucleic acid molecules of claims 3-5, said method comprising contacting said nucleic acid molecule with an agent under conditions that allow the agent to modulate the level or activity of the nucleic acid molecule, wherein said modulation is in a tissue selected from the group consisting of mobilized peripheral blood $CD34^+$ cells, in HL60 (promyelocytic leukemia) cell line, $CD34^-$ mobilized bone marrow cells, $CD8^+$T lymphocytes, $CD34^+$ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, $CD34^-$ mobilized peripheral blood cells, and $CD4^+$T lymphocytes.

48. The method of claim 31 wherein said detecting is in a cell derived from a subject having a disorder involving said cell.

49. The method of claim 47 wherein said modulation is in a subject having a disorder involving said cell.

50. The method of claim 48 wherein said disorder is anemia, neutropenia, or thrombocytopenia.

51. The method of claim 49 wherein said disorder is anemia, neutropenia, or thrombocytopenia.

52. The method of claim49, wherein said disorder involves inflammation.

Input file Flhb14275cons; Output File 14275tra
Sequence length 2100

FIG. 1A.

```
CGCGTCCGCTGAGCCCTCACGGGACATCTGTGCCCCTCATGGGACACCTGTGTCCTCACAGTACACTTGTGACCCTTCC

AGGACACCTTACTGGTAGAATTAGTGTAGCTGCCCCCACCCTGAGGCCAAGGACACCATTGTCTCAGGAAGGCTGAAGA

CCACAGGCTCCTGGGGGGACAGAGGGCAGGTGGGGCCCCTCAGGACCCTCCTTGGTGGAAACCAAGACCAGCAAGGCGG

                                                M  N  A  T  G  T  P  V          8
GTGGCTCCACCCTGCGTCGGGCCTCAGTCAGCCCCCGGGGGAGGCC ATG AAC GCC ACG GGG ACC CCG GTG    24

 A  P  E  S  C  Q  Q  L  A  A  G  G  H  S  R  L  I  V  L  H                    28
GCC CCC GAG TCC TGC CAA CAG CTG GCG GCC GGC GGG CAC AGC CGG CTC ATT GTT CTG CAC    84

 Y  N  H  S  G  R  L  A  G  R  G  G  P  E  D  G  G  L  G  A                    48
TAC AAC CAC TCG GGC CGG CTG GCC GGG CGC GGG GGG CCG GAG GAT GGC GGC CTG GGG GCC   144

 L  R  G  L  S  V  A  A  S  C  L  V  V  L  E  N  L  L  V  L                    68
CTG CGG GGG CTG TCG GTG GCC GCC AGC TGC CTG GTG GTG CTG GAG AAC TTG CTG GTG CTG   204

 A  A  I  T  S  H  M  R  S  R  R  W  V  Y  Y  C  L  V  N  I                    88
GCG GCC ATC ACC AGC CAC ATG CGG TCG CGA CGC TGG GTC TAC TAT TGC CTG GTG AAC ATC   264

 T  L  S  D  L  L  T  G  A  A  Y  L  A  N  V  L  L  S  G  A                   108
ACG CTG AGT GAC CTG CTC ACG GGC GCG GCC TAC CTG GCC AAC GTG CTG CTG TCG GGG GCC   324

 R  T  F  R  L  A  P  A  Q  W  F  L  R  E  G  L  L  F  T  A                   128
CGC ACC TTC CGT CTG GCG CCC GCC CAG TGG TTC CTA CGG GAG GGC CTG CTC TTC ACC GCC   384

 L  A  L  S  T  F  S  L  L  F  T  A  G  E  R  F  A  T  M  V                   148
CTG GCC GCC TCC ACC TTC AGC CTG CTC TTC ACT GCA GGG GAG CGC TTT GCC ACC ATG GTG   444

 R  P  V  A  E  S  G  A  T  K  T  S  R  V  Y  G  F  I  G  L                   168
CGG CCG GTG GCC GAG AGC GGG GCC AAG ACC AGC CGC GTC TAC GGC TTC ATC GGC CTC   504

 C  W  L  L  A  A  L  L  G  M  L  P  L  L  G  W  N  C  L  C                   188
TGC TGG CTG CTG GCC GCG CTG CTG GGG ATG CTG CCT TTG CTG GGC TGG AAC TGC CTG TGC   564

 A  F  D  R  C  S  S  L  L  P  L  Y  S  K  R  Y  I  L  F  C                   208
GCC TTT GAC CGC TGC TCC AGC CTT CTG CCC CTC TAC TCC AAG CGC TAC ATC CTC TTC TGC   624

 L  V  I  F  A  G  V  L  A  T  I  M  G  L  Y  G  A  I  F  R                   228
CTG GTG ATC TTC GCC GGC GTC CTG GCC ACC ATC ATG GGC CTC TAT GGG GCC ATC TTC CGC   684

 L  V  Q  A  S  G  Q  K  A  P  R  P  A  A  R  R  K  A  R  R                   248
CTG GTG CAG GCC AGC GGG CAG AAG GCC CCA CGC CCA GCG GCC CGC CGC AAG GCC CGC CGC   744

 L  L  K  T  V  L  M  I  L  L  A  F  L  V  C  W  G  P  L  F                   268
CTG CTG AAG ACG GTG CTG ATG ATC CTG CTG GCC TTC CTG GTG TGC TGG GGC CCA CTC TTC   804

 G  L  L  L  A  D  V  F  G  S  N  L  W  A  Q  E  Y  L  R  G                   288
GGG CTG CTG CTG GCC GAC GTC TTT GGC TCC AAC CTC TGG GCC CAG GAG TAC CTG CGG GGC   864

 M  D  W  I  L  A  L  A  V  L  N  S  A  V  N  P  I  I  Y  S                   308
ATG GAC TGG ATC CTG GCC CTG GCC GTC CTC AAC TCG GCG GTC AAC CCC ATC ATC TAC TCC   924

 F  R  S  R  E  V  C  P  A  V  L  S  F  L  C  C  G  C  L  R                   328
TTC CGC AGC AGG GAG GTG TGC AGA GCC GTG CTC AGC TTC CTC TGC TGC GGG TGT CTC CGG   984

 L  G  M  R  G  P  G  D  C  L  A  R  A  V  E  A  H  S  G  A                   348
CTG GGC ATG CGA GGG CCC GGG GAC TGC CTG GCC CGG GCC GTC GAG GCT CAC TCC GGA GCT  1044

 S  T  T  D  S  S  L  R  P  R  D  S  F  R  G  S  R  S  L  S                   368
```

TO FIG. 1B.

FROM FIG. 1A.

```
TCC ACC ACC GAC AGC TCT CTG AGG CCA AGG GAC AGC TTT CGC GGC TCC CGC TCG CTC AGC  1104
 F   R   M   R   E   P   L   S   S   I   S   S   V   R   S   I   *               385
TTT CGG ATG CGG GAG CCC CTG TCC AGC ATC TCC AGC GTG CGG AGC ATC TGA              1155

AGTTGCAGTCTTGCGTGTGGATGGTGCAGCCACCGGGTGCGTGCCAGGCAGGCCCTCCTGGGGTACAGGAAGCTGTGTG
CACGCAGCCTCGCCTGTATGGGGAGCAGGGAACGGGACAGGCCCCCATGGTCTTCCCGGTGGCCTCTCGGGGCTTCTGA
CGCCAAATGGGCTTCCCATGGTCACCCTGGACAAGGAGGCAACCACCCCACCTCCCCGTAGGAGCAGAGAGCACCCTGG
TGTGGGGGCGAGTGGGTTCCCCACAACCCCGCTTCTGTGTGATTCTGGGGAAGTCCCGGCCCCTCTCTGGGCCTCAGTA
GGGCTCCCAGGCTGCAAGGGGTGGACTGTGGGATGCATGCCCTGGCAACATTGAAGTTCGATCATGGTAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

FIG. 1B.

Prosite Pattern Matches for Flh14275orfaa

>PS00001/PDOC00001/ASN-GLYCOSYLATION N-glycosylation site.

Query:	2	NATG	5
Query:	30	NHSG	33
Query:	87	NITL	90

>PS00005/PDOC00005/PDC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query:	32	SGR	34
Query:	77	SRR	79
Query:	110	TFR	112
Query:	159	TSR	161
Query:	201	SKR	203
Query:	308	SFR	310
Query:	354	SLR	356
Query:	360	SFR	362
Query:	368	SFR	370
Query:	380	SVR	382

>PS00006/PDOC00006/CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query:	89	TLSD	92
Query:	139	TAGE	142
Query:	349	STTD	352

>ps00008/pdoc00008/myristyl n-MYRISTOYLATION SITE.

Query:	44	GGLGAL	49
Query:	51	GLSVAA	56
Query:	123	GLLFTA	128
Query:	155	GATKTS	160
Query:	214	GVLATI	219
Query:	221	GLYGAI	226
Query:	269	GLLLAD	274
Query:	347	GASTTD	352

FIG. 2.

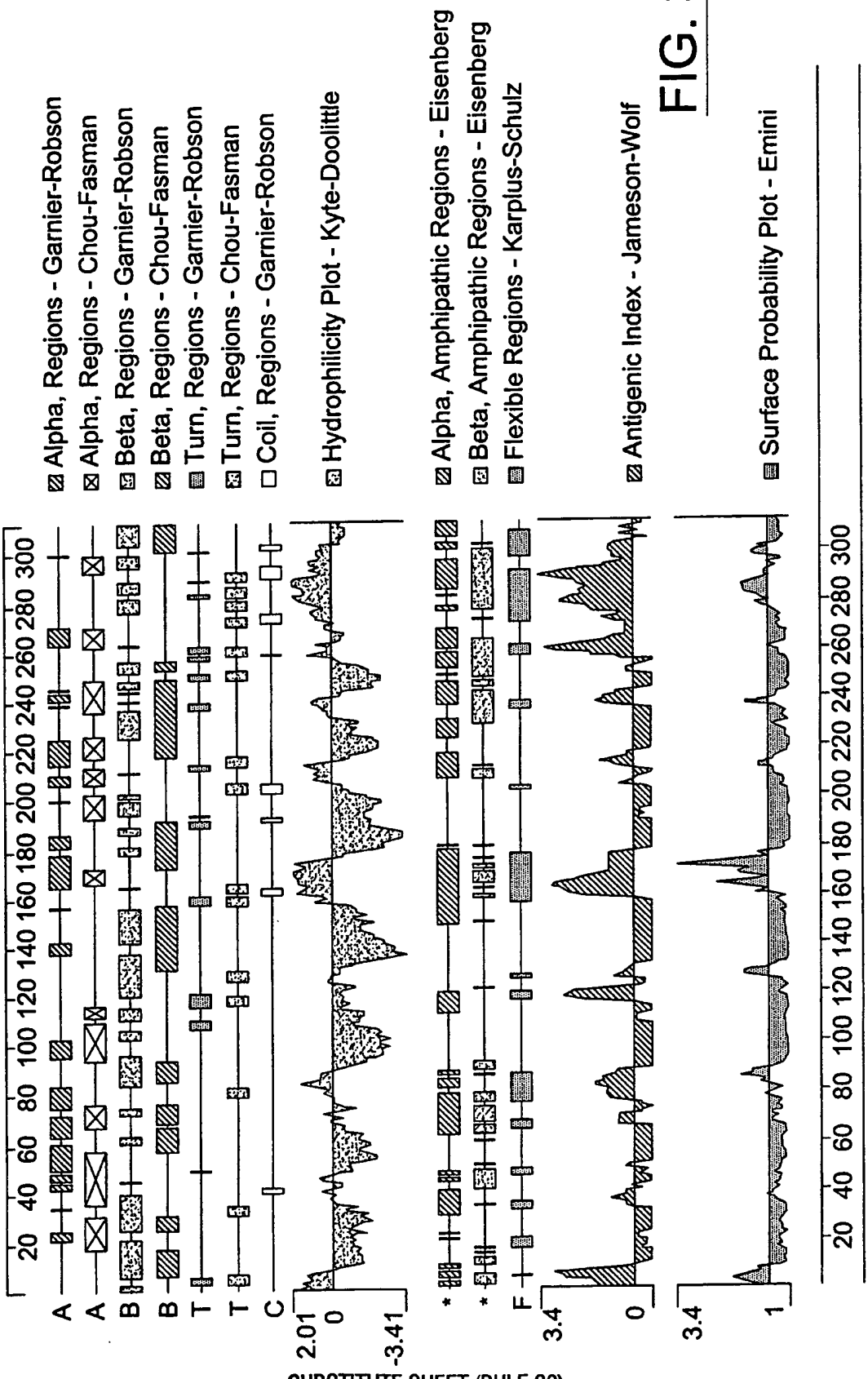
Alpha, Regions - Garnier-Robson
Alpha, Regions - Chou-Fasman
Beta, Regions - Garnier-Robson
Beta, Regions - Chou-Fasman
Turn, Regions - Garnier-Robson
Turn, Regions - Chou-Fasman
Coil, Regions - Garnier-Robson
Hydrophilicity Plot - Kyte-Doolittle
Alpha, Amphipathic Regions - Eisenberg
Beta, Amphipathic Regions - Eisenberg
Flexible Regions - Karplus-Schulz
Antigenic Index - Jameson-Wolf
Surface Probability Plot - Emini
20 40 60 80 100 120 140 160 180 200 220 240 260 280 300
A
A
B
B
T
T
C
2.01
0
-3.41
*
*
F
3.4
0
3.4
1
20 40 60 80 100 120 140 160 180 200 220 240 260 280 300

FIG. 3.

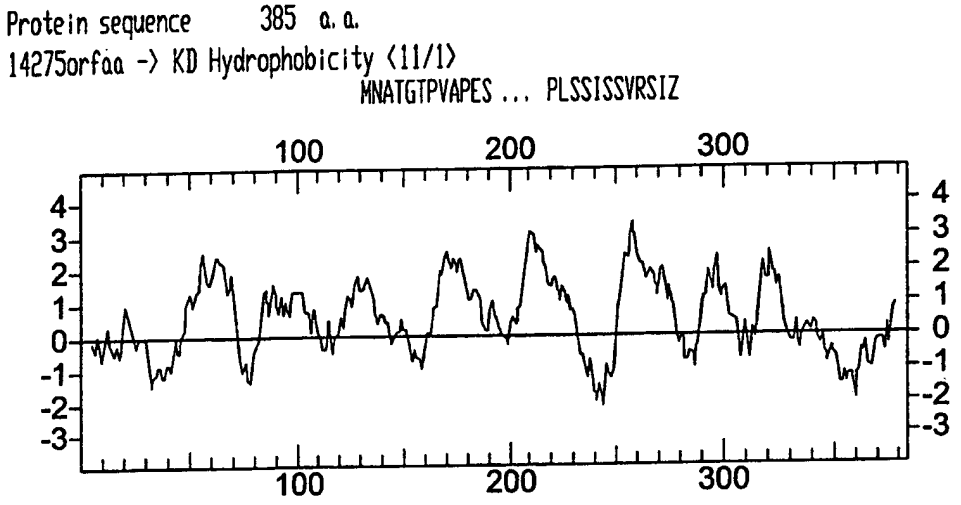
Protein sequence 385 a.a.
14275orfaa -> KD Hydrophobicity (11/1)
MNATGTPVAPES ... PLSSISSVRSIZ
100
200
300
4
3
2
1
0
-1
-2
-3

FIG. 4.

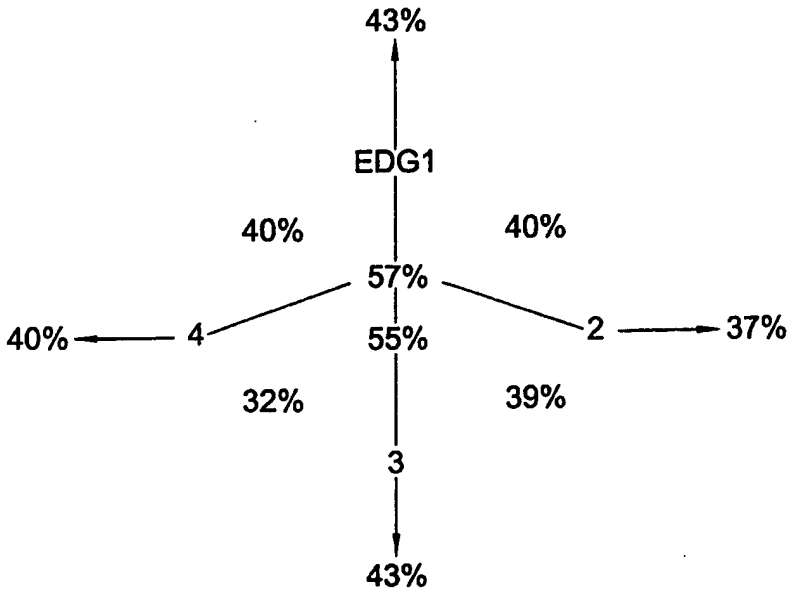
43%
EDG1
40%
40%
57%
40%
4
55%
2
37%
32%
39%
3
43%

FIG. 5.

Protein Family/Domain HMM Matches for Flh14275orfaa

>PF00001/7tm_1 7 transmembrane receptor (rhodopsin family)

Score: 135.54 Seq: 63 307 Model: 1 269

```
                 =GNiLVIWvIcRyRRMRTPMNYFIvNLAvADLLFslftMPFWMvYyvMqg
                  +N+LV+++I+ ++R R+  +Y +VN  ++DLL + + +   +++ +
Flh14275or  63    ENLLVLAAITSHMRSRRWVYYCLVNITLSDLLTGAAYLANVLLSGAR-T  110

                 RWpFGdfMCrIWmYFDYMNMYASIFFLTcISIDRYLWAICHPMrYmRWMT
                  + + ++ + +  ++ ++++ AS F+L++ + +R+  ++++P+   + T
Flh14275or 111    -FRLAPAQWFLREGLLFTALAASTFSLLFTAGERF-ATMVRPVA-ESGAT  157

                 pRHRAWvMIiiIWvMSFlISMPPFLMFrWstyrDEneWNmTWCmIyDWPe
                 ++ R++  I+++W++++L++M+P LL++ W ++    +  +  C+   P
Flh14275or 158    KTSRVYGFIGLCWLLAALLGMLP-LLG-W-NCLCAFD--R--CSS-LLPL  199

                 wMWrWYvILmtiimgFYIPMiIMlFCYwRIYRIaRlWMRMIpswQrRRrm
                 ++ R ++ + ++ + F   +  ++ +Y+ I+R    +++ +   +  R+
Flh14275or 200    YSKRYEL-FCLV-IFAGVLATIMGLYGAIFRL--VQASGQ---KAPRPA  241

                 SmRrERRivKMliIIMvVFIICWlPYFIvmfMDTLMMwwFCefC. Iwrrl
                   R+ RR++K++++I+++F++CW+P+F +++        +F + +   + L
Flh14275or 242    ARRKARRLLKTVLMILLAFLVCWGPLFGLLL-----ADVFGSNLWAQEYL  286

                 WmY. IfeWLaYvNCpCiNPIIY*
                    +++ LA +N++ +NPIIY
Flh14275or 287    RGMDWILALAVLNSA-VNPIIY  307
```

FIG. 6.

SEQUENCE LISTING

<110> Glucksmann, Maria A.

<120> 14275 Receptor, A Novel G-Protein Coupled Receptor
Related To The EDG Receptor Family

<130> 035800/169287

<140> 09/146,416
<141> 1998-09-03

<160> 3

<170> PatentIn Ver. 2.0

<210> 1
<211> 384
<212> PRT
<213> Homo sapiens

<400> 1

```
Met Asn Ala Thr Gly Thr Pro Val Ala Pro Glu Ser Cys Gln Gln Leu
  1               5                  10                  15

Ala Ala Gly Gly His Ser Arg Leu Ile Val Leu His Tyr Asn His Ser
             20                  25                  30

Gly Arg Leu Ala Gly Arg Gly Gly Pro Glu Asp Gly Gly Leu Gly Ala
         35                  40                  45

Leu Arg Gly Leu Ser Val Ala Ala Ser Cys Leu Val Val Leu Glu Asn
     50                  55                  60

Leu Leu Val Leu Ala Ala Ile Thr Ser His Met Arg Ser Arg Arg Trp
 65                  70                  75                  80

Val Tyr Tyr Cys Leu Val Asn Ile Thr Leu Ser Asp Leu Leu Thr Gly
                 85                  90                  95

Ala Ala Tyr Leu Ala Asn Val Leu Leu Ser Gly Ala Arg Thr Phe Arg
            100                 105                 110

Leu Ala Pro Ala Gln Trp Phe Leu Arg Glu Gly Leu Leu Phe Thr Ala
        115                 120                 125

Leu Ala Ala Ser Thr Phe Ser Leu Leu Phe Thr Ala Gly Glu Arg Phe
    130                 135                 140
```

```
Ala Thr Met Val Arg Pro Val Ala Glu Ser Gly Ala Thr Lys Thr Ser
145                 150                 155                 160

Arg Val Tyr Gly Phe Ile Gly Leu Cys Trp Leu Leu Ala Ala Leu Leu
                165                 170                 175

Gly Met Leu Pro Leu Leu Gly Trp Asn Cys Leu Cys Ala Phe Asp Arg
            180                 185                 190

Cys Ser Ser Leu Leu Pro Leu Tyr Ser Lys Arg Tyr Ile Leu Phe Cys
        195                 200                 205

Leu Val Ile Phe Ala Gly Val Leu Ala Thr Ile Met Gly Leu Tyr Gly
    210                 215                 220

Ala Ile Phe Arg Leu Val Gln Ala Ser Gly Gln Lys Ala Pro Arg Pro
225                 230                 235                 240

Ala Ala Arg Arg Lys Ala Arg Arg Leu Leu Lys Thr Val Leu Met Ile
                245                 250                 255

Leu Leu Ala Phe Leu Val Cys Trp Gly Pro Leu Phe Gly Leu Leu Leu
            260                 265                 270

Ala Asp Val Phe Gly Ser Asn Leu Trp Ala Gln Glu Tyr Leu Arg Gly
        275                 280                 285

Met Asp Trp Ile Leu Ala Leu Ala Val Leu Asn Ser Ala Val Asn Pro
    290                 295                 300

Ile Ile Tyr Ser Phe Arg Ser Arg Glu Val Cys Arg Ala Val Leu Ser
305                 310                 315                 320

Phe Leu Cys Cys Gly Cys Leu Arg Leu Gly Met Arg Gly Pro Gly Asp
                325                 330                 335

Cys Leu Ala Arg Ala Val Glu Ala His Ser Gly Ala Ser Thr Thr Asp
            340                 345                 350

Ser Ser Leu Arg Pro Arg Asp Ser Phe Arg Gly Ser Arg Ser Leu Ser
        355                 360                 365

Phe Arg Met Arg Glu Pro Leu Ser Ser Ile Ser Ser Val Arg Ser Ile
    370                 375                 380
```

<210> 2
<211> 1877
<212> DNA
<213> Homo sapiens

<400> 2
cgcgtccgct gagccctcac gggacatctg tgcccctcat gggacacctg tgtcctcaca 60
gtacacttgt gacccttcca ggacacctta ctggtagaat tagtgtagct gcccccaccc 120
tgaggccaag gacaccattg tctcaggaag gctgaagacc acaggctcct ggggggacag 180
agggcaggtg gggcccctca ggaccctcct tggtggaaac caagaccagc aaggcgggtg 240
gctccaccct gcgtcgggcc tcagtcagcc cccgggggag gccatgaacg ccacggggac 300
cccggtggcc cccgagtcct gccaacagct ggcggccggc gggcacagcc ggctcattgt 360
tctgcactac aaccactcgg gccggctggc cgggcgcggg gggccggagg atggcggcct 420
gggggccctg cgggggctgt cggtggccgc cagctgcctg gtggtgctgg agaacttgct 480
ggtgctggcg gccatcacca gccacatgcg gtcgcgacgc tgggtctact attgcctggt 540
gaacatcacg ctgagtgacc tgctcacggg cgcggcctac ctggccaacg tgctgctgtc 600
gggggcccgc accttccgtc tggcgcccgc ccagtggttc ctacgggagg gcctgctctt 660
caccgccctg gccgcctcca ccttcagcct gctcttcact gcaggggagc gctttgccac 720
catggtgcgg ccggtggccg agagcggggc caccaagacc agccgcgtct acggcttcat 780
cggcctctgc tggctgctgg ccgcgctgct ggggatgctg cctttgctgg gctggaactg 840
cctgtgcgcc tttgaccgct gctccagcct tctgcccctc tactccaagc gctacatcct 900
cttctgcctg gtgatcttcg ccggcgtcct ggccaccatc atgggcctct atggggccat 960
cttccgcctg gtgcaggcca gcgggcagaa ggccccacgc ccagcggccc gccgcaaggc 1020
ccgccgcctg ctgaagacgg tgctgatgat cctgctggcc ttcctggtgt gctggggccc 1080
actcttcggg ctgctgctgg ccgacgtctt tggctccaac ctctgggccc aggagtacct 1140
gcggggcatg gactggatcc tggccctggc cgtcctcaac tcggcggtca accccatcat 1200
ctactccttc cgcagcaggg aggtgtgcag agccgtgctc agcttcctct gctgcgggtg 1260
tctccggctg ggcatgcgag ggcccgggga ctgcctggcc cgggccgtcg aggctcactc 1320
cggagcttcc accaccgaca gctctctgag gccaagggac agctttcgcg gctcccgctc 1380
gctcagcttt cggatgcggg agcccctgtc cagcatctcc agcgtgcgga gcatctgaag 1440
ttgcagtctt gcgtgtggat ggtgcagcca ccgggtgcgt gccaggcagg ccctcctggg 1500
gtacaggaag ctgtgtgcac gcagcctcgc ctgtatgggg agcagggaac gggacaggcc 1560
cccatggtct tcccggtggc ctctcggggc ttctgacgcc aaatgggctt cccatggtca 1620
ccctggacaa ggaggcaacc accccacctc cccgtaggag cagagagcac cctggtgtgg 1680
gggcgagtgg gttccccaca accccgcttc tgtgtgattc tggggaagtc ccggcccctc 1740
tctgggcctc agtagggctc ccaggctgca aggggtggac tgtgggatgc atgccctggc 1800
aacattgaag ttcgatcatg gtaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1860
aaaaaaaaaa aaaaaaa 1877

<210> 3
<211> 269
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown Organism: Transmembrane

Receptor of the Rhodopsin Superfamily

<400> 3

```
Gly Asn Ile Leu Val Ile Trp Val Ile Cys Arg Tyr Arg Arg Met Arg
 1               5                  10                  15

Thr Pro Met Asn Tyr Phe Ile Val Asn Leu Ala Val Ala Asp Leu Leu
            20                  25                  30

Phe Ser Leu Phe Thr Met Pro Phe Trp Met Val Tyr Tyr Val Met Gln
        35                  40                  45

Gly Arg Trp Pro Phe Gly Asp Phe Met Cys Arg Ile Trp Met Tyr Phe
    50                  55                  60

Asp Tyr Met Asn Met Tyr Ala Ser Ile Phe Phe Leu Thr Cys Ile Ser
65                  70                  75                  80

Ile Asp Arg Tyr Leu Trp Ala Ile Cys His Pro Met Arg Tyr Met Arg
                85                  90                  95

Trp Met Thr Pro Arg His Arg Ala Trp Val Met Ile Ile Ile Ile Trp
            100                 105                 110

Val Met Ser Phe Leu Ile Ser Met Pro Pro Phe Leu Met Phe Arg Trp
        115                 120                 125

Ser Thr Arg Tyr Asp Glu Asn Glu Trp Asn Met Thr Trp Cys Met Ile
    130                 135                 140

Tyr Asp Trp Pro Glu Trp Met Trp Arg Trp Tyr Val Ile Leu Met Thr
145                 150                 155                 160

Ile Ile Met Gly Phe Tyr Ile Pro Met Ile Ile Met Leu Phe Cys Tyr
                165                 170                 175

Trp Arg Ile Tyr Arg Ile Ala Arg Leu Trp Met Arg Met Ile Pro Ser
            180                 185                 190

Trp Gln Arg Arg Arg Arg Met Ser Met Arg Arg Glu Arg Arg Ile Val
        195                 200                 205

Lys Met Leu Ile Ile Ile Met Val Val Phe Ile Ile Cys Trp Leu Pro
    210                 215                 220

Tyr Phe Ile Val Met Phe Met Asp Thr Leu Met Met Trp Trp Phe Cys
225                 230                 235                 240
```

Glu Phe Cys Ile Trp Arg Arg Leu Trp Met Tyr Ile Phe Glu Trp Leu
245 250 255

Ala Tyr Val Asn Cys Pro Cys Ile Asn Pro Ile Thr Tyr
260 265

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/20347

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12N 15/12; C12P 21/02; C07K 14/47, 14/705

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 71.1, 325, 471, 320.1; 530/350; 536/23.5, 24.3, 24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

West (US Patent full), STN via medline, caplus, embase, search terms: G-protein coupled receptors, EDG and lysophosphatidic acid, 14725 receptor. Sequence search for SEQ ID NOs:1 and 2.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LEE et al. Lysophosphatidic acid stimulates the G-protein-coupled receptor EDG-1 as a low affinity agonist. Journal of Biological Chemistry. 21 August 1998, Vol. 273, No. 34, pages 22105-22112, see entire document.	1, 3-8
A	FUKUSHIMA et al. A single receptor encoded by vzg-1/lpA1/edg-2 couples to G proteins and mediates multiple cellular responses to lysophosphatidic acid. Proceedings of the National Academy of Sciences of the United States of America. May 1998, Vol. 95, No. 11, pages 6151-6156, see entire document.	1, 3-8

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 21 DECEMBER 1999	Date of mailing of the international search report 02 FEB 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer FOZIA HAMUD Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/20347

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GRALER et al. EDG6, a novel G-protein-coupled receptor related to receptors for bioactive lysophospholipids, is specifically expressed in lymphoid tissue. Genomics. October 1998, Vol.53, pages 164-169, see entire document.	1, 3-8
A	ZONDAG et al. Sphingosine 1-phosphate signalling through the G-protein-coupled receptor edg-1. Biochemical Journal. March 1998, Vol. 330, No.2, pages 605-609, see entire document.	1, 3-8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/20347

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1, 3-8

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/20347

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/69.1, 71.1, 325, 471, 320.1; 530/350; 536/23.5, 24.3, 24.31

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.
Group I, claims 1, 3-8, drawn to an isolated nucleic acid molecule encoding a G-protein coupled receptor, a vector, a host cell, a method of producing the encoded polypeptide, and the encoded G-protein coupled receptor.
Group II, claim 2, drawn to an isolated antibody that selectively binds to the receptor.
Group III, claims 9-37, drawn to a method for identifying an agent that binds, interacts or modulates the level or activity of the G-protein coupled receptor, comprising, contacting the agent with a cell capable of expressing the receptor, and measuring the level or activity of said receptor.
Group IV, claims 38-52, drawn to a method for identifying an agent that modulates the level or activity of the nucleic acid that encodes the G-protein coupled receptor, comprising, contacting the agent with a cell capable of expressing said nucleic acid molecule, and measuring the level or activity of said nucleic acid molecule.
The inventions listed as Groups I-VII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Pursuant to 37 C.F.R § 1.475 (d), the ISA/US considers that where multiple products and processes are claimed, the main invention shall consist of the first invention of the category first mentioned in the claims and the first recited invention of each of the other categories related thereto. Accordingly, the main invention (Group I)comprises the first-recited product, the isolated polypeptide and the nucleic acid molecule encoding it, a vector comprising said nucleic acid molecule, a host cell and a method of producing the encoded polypeptide. Further pursuant to 37 C.F.R § 1.475 (d), the ISA/US considers that any feature which the subsequently recited products and methods share with the main invention does not constitute a special technical feature within the meaning of PCT Rule 13.2 and that each of such products and methods accordingly defines a separate invention.